CT biomarkers in lung cancer screening

Xie, Xueqian

CT Biomarkers in Lung Cancer Screening

Xueqian Xie

Xueqian Xie 解学乾
CT biomarkers in lung cancer screening
PhD thesis University of Groningen, with a summary in Dutch

Cover: Birds in the Pomegranate, designed by Xueqian Xie
Layout: Xueqian Xie

The publication of this thesis was financially supported by
University of Groningen, University Medical Center Groningen

RIJKSUNIVERSITEIT GRONINGEN

CT Biomarkers in Lung Cancer Screening

Proefschrift

ter verkrijging van het doctoraat in de
Medische Wetenschappen
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. E. Sterken,
in het openbaar te verdedigen op
maandag 16 september 2013
om 16:15 uur

door

Xueqian Xie

geboren op 6 februari 1977

te Liaoning, China

Promotor : Prof. dr. M. Oudkerk

Copromotor : Dr. R. Vliegenthart

Beoordelingscommissie : Prof. dr. J.W.J. Lammers
Prof. dr. W.P.Th.M. Mali
Prof. dr. H.J.M. Groen

ISBN: 978-90-367-6293-9 (book)
978-90-367-6292-2 (e-book)

List of contents

Part 4 Emphysema and bronchial wall

Chapter 1

General Introduction

Quantitative imaging biomarkers gain increasing importance in research and clinical medicine. Measurement of biomarkers compared to mere visual evaluation allows for more objective evaluation of disease, and more accurate assessment of the stage and development of disease through time [1-6]. Low-dose computed tomography (CT) examinations, obtained as part of lung cancer screening, can provide quantitative information on biomarkers for lung cancer, cardiovascular disease and chronic obstructive pulmonary disease (COPD) - the big three killers of the Western population [7-12].

The Dutch-Belgian Randomized Lung Cancer Screening Trial (Dutch acronym: NELSON) was launched in 2003. Details on participant recruitment and CT acquisition protocol were described elsewhere [8]. In short, four rounds of low-dose CT screening took place in the year 1, 2, 4, and 6 for heavy (ex-) smokers between 50 and 75 years of age. Evaluation of CT examinations in lung cancer screening can also include assessment of coronary calcification, emphysema and airway wall, markers for abovementioned major diseases that share risk factors with lung cancer [13]. A practical approach to the radiological evaluation of CT lung cancer screening examinations is necessary, including evaluation of pulmonary nodules and non-nodular diseases.

Lung nodule volumetry

The most common cause of cancer-related death is lung cancer. In the year 2000, lung cancer accounted for 17% of total cancer mortality [14]. Low-dose CT was proposed as a promising method to screen for lung cancer. Several cohort studies and randomised clinical trials using low-dose CT were started, aiming to investigate the effect of lung cancer screening on the distribution of tumour stages, and eventually the effect on survival [7, 15, 16]. In CT screening, an overwhelming number of small, indeterminate solid pulmonary nodules are detected [7, 9, 17]. The large majority of these nodules are benign. Optimal management protocol of small pulmonary nodules is important for early identification of malignant lesions in screening, while at the same time preventing too many false positive referrals. Efficient lung cancer screening starts with sensitive observer detection of pulmonary nodules. There are only scarce validation data on the detectability of small pulmonary nodules by low-dose CT [18, 19].

In lung cancer screening, treatment decisions usually depend on pulmonary nodule size at first detection, and on growth rate at follow-up [9]. Therefore, it is essential to assess nodule size and growth rate accurately and reproducibly [20]. Unfortunately, variability has been found in CT-derived nodule size assessment [21, 22]. In view of the current practice that patients frequently undergo follow-up examinations, sometimes on different scanners, reliable inter- and intra-scanner reproducibility of nodule volumetry is important. However, previous studies reported inconsistent results regarding the reproducibility of nodule volumetry. Some in-vitro studies have been performed in which artificial nodules

were placed at known locations in a thoracic phantom without pulmonary vessels [23-25]. Some of these studies were based on older generation CT scanners [23, 26]. These studies generally showed a small margin of variability in nodule volumetry for software from different vendors. On the other side, in-vivo studies have shown that variability can be considerable, with variability up to 25% for 15 to 500 mm^3 nodules [21, 27-29]. A new anthropomorphic phantom with pulmonary background is ideal to validate the detectability and volumetry of nodules, as in this controlled setting, a known number of artificial nodules are inserted, all having a known volume. These validation studies are essential for optimization of the management protocol for small nodules in lung cancer screening [8, 30].

Coronary calcium scoring

The CT-derived coronary artery calcium score is a strong predictive biomarker of cardiovascular events [31-34]. Calcium scoring has been found to improve cardiovascular risk stratification beyond cardiovascular risk factors [33, 35]. Due to the irregular and periodic movements of coronary arteries, electrocardiography (ECG)-triggered cardiac acquisition techniques are applied in CT to minimize motion artifacts and optimize calcium scoring [32]. Compared to ECG-triggered CT, nontriggered CT of the thorax is extensively utilized. In 2007, 13.6 million nontriggered thoracic CT examinations were performed in the United States alone, in contrast to 0.7 million ECG-triggered CT examinations for calcium scoring [36]. In lung cancer screening, coronary calcification is a frequent finding [11]. Age and smoking, the current selection criteria for lung cancer screening, are also correlated with coronary calcification and coronary heart disease [37]. If nontriggered CT can be used for calcium scoring, to stratify individuals in categories of cardiovascular risk and to identify those at high cardiovascular risk, there may be a substantial unused primary prevention potential [38]. Also, deriving the calcium score from the same examination as used in lung cancer screening may positively impact the cost-effectiveness of screening. Because motion of coronary arteries influences calcium scoring [39], the utilization of coronary calcium scoring in nontriggered CT is still being debated [40]. Variability in coronary calcium quantification between ECG-triggered and nontriggered CT can be substantial [39, 41]. Compared with the extensive publications regarding ECG-triggered cardiac CT, the literature on calcium scoring based on nontriggered thoracic CT is relatively limited.

Emphysema and bronchial wall quantification

COPD is characterized by airflow limitation that is not fully reversible [42]. The pathogenesis of airflow limitation in COPD is mainly related to emphysema and small airway remodelling [43]. For COPD patients, quantitative morphological biomarkers based on CT could be important to understand pathogenesis and the effect of therapeutic interventions [44, 45], and could help to identify those most at risk for acute exacerbations [46]. In mul-

tiple studies, multi-detector CT was found to accurately evaluate the extent of emphysema [47-50]. Different measures have been introduced to quantify the severity of emphysema [51-55]. Airway wall quantification started over a decade ago, mainly for large airways [44, 56-58], but investigators have measured peripheral airways down to 0.5 mm lumen diameter [59] and 2.8 mm outer diameter [60]. Measurement of narrowing of CT-detectable airways may estimate the degree of small airway disease [59]. The assumption is that emphysema and peripheral airway wall thickness, as CT-derived biomarkers, are correlated to airflow obstruction in COPD patients. Study results have been variable and sometimes conflicting. However, some individual studies have been small and underpowered [43, 61-64].

Nearly half of the smokers in lung cancer screening have chronic respiratory symptoms, i.e., chronic hyper secretion of mucus, combined with chronic cough, often accompanied by dyspnea and wheezing [65]. Smokers with these symptoms have a higher risk to develop chronic bronchitis, a disease associated with an accelerated decline in pulmonary function – a major risk factor for COPD and all-cause mortality [66, 67]. Despite the high prevalence in smokers, chronic bronchitis is often under-diagnosed or diagnosed late [68]. Histopathologically, chronic bronchitis characterized by bronchial wall thickening and airway luminal narrowing, commonly in large airways [69, 70]. Morphological biomarkers are important to understand pathogenesis and effect of therapeutic interventions for chronic bronchitis [71, 72]. In participants of lung cancer screening, it is important to know whether morphological changes of airway walls are related to chronic respiratory symptoms, since early detection of chronic bronchitis by CT could lead to earlier treatment.

Outline thesis

Quantitative CT biomarkers for lung cancer, cardiovascular disease and COPD were investigated in this thesis, in particular quantification of the pulmonary nodule, coronary calcification, emphysema and the bronchial wall. In **Chapter 2**, a practical approach to evaluate these CT biomarkers in the NELSON trial is described. **Chapter 3 - 5** describe studies on lung nodule validation based on an anthropomorphic phantom and on screening data. Here the focus is on the sensitivity of detection of nodules in low-dose CT, and the accuracy and variability of nodule volumetry by manual and semi-automated measurements. In **Chapter 6**, the diagnostic and prognostic significance of coronary calcification assessed on nontriggered thoracic CT is described in a systematic review and meta-analysis, while **Chapter 7** evaluates the agreement in coronary calcium scoring between thoracic CT and dedicated cardiac CT. In **Chapter 8**, emphysema and airway measurements based on CT are correlated with airflow obstruction in COPD in a systematic review and meta-analysis. Finally, in **Chapter 9**, CT-derived airway wall thickness along the respiratory pathway is compared between individuals with and without chronic respiratory symptoms.

References

1. Dienstmann R, Rodon J, Tabernero J. Biomarker-driven patient selection for early clinical trials. Current opinion in oncology 2013; 25:305-312
2. Tsai CL, Koong AC, Hsu FM, Graber M, Chen IS, Cheng JC. Biomarker studies on radiotherapy to hepatocellular carcinoma. Oncology 2013; 84 Suppl 1:64-68
3. Sherwi N, Merali S, Wong K. Personalizing biomarker strategies in heart failure with galectin-3. Future cardiology 2012; 8:885-894
4. Nowosinska E, Avril S, Murray I, Szyszko T, Avril N. FDG-PET/CT as a molecular biomarker in ovarian cancer. Cancer biomarkers : section A of Disease markers 2010; 8:167-175
5. Buckler AJ, Mulshine JL, Gottlieb R, Zhao B, Mozley PD, Schwartz L. The use of volumetric CT as an imaging biomarker in lung cancer. Acad Radiol 2010; 17:100-106
6. Pedrosa I, Alsop DC, Rofsky NM. Magnetic resonance imaging as a biomarker in renal cell carcinoma. Cancer 2009; 115:2334-2345
7. Aberle DR, Adams AM, Berg CD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med 2011; 365:395-409
8. Ru Zhao Y, Xie X, de Koning HJ, Mali WP, Vliegenthart R, Oudkerk M. NELSON lung cancer screening study. Cancer imaging 2011; 11 S79-84
9. van Klaveren RJ, Oudkerk M, Prokop M, et al. Management of lung nodules detected by volume CT scanning. The New England journal of medicine 2009; 361:2221-2229
10. Jacobs PC, Gondrie MJ, Mali WP, et al. Unrequested information from routine diagnostic chest CT predicts future cardiovascular events. European radiology 2011; 21:1577-1585
11. Jacobs PC, Gondrie MJA, van der Graaf Y, et al. Coronary artery calcium can predict all-cause mortality and cardiovascular events on low-dose CT screening for lung cancer. AJR American journal of roentgenology 2012; 198:505-511
12. Shemesh J, Henschke CI, Shaham D, et al. Ordinal scoring of coronary artery calcifications on low-dose CT scans of the chest is predictive of death from cardiovascular disease. Radiology 2010; 257:541-548
13. Lozano R, Naghavi M, Foreman K, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2013; 380:2095-2128
14. Shibuya K, Mathers CD, Boschi-Pinto C, Lopez AD, Murray CJ. Global and regional estimates of cancer mortality and incidence by site: II. Results for the global burden of disease 2000. BMC Cancer 2002; 2:37
15. Gatsonis CA, Natl Lung Screening Trial Res Team. The National Lung Screening Trial: overview and study design. Radiology 2011; 258:243-253

16. Nair A, Hansell DM. European and North American lung cancer screening experience and implications for pulmonary nodule management. Eur Radiol 2011; 21:2445-2454
17. Xu DM, van Klaveren RJ, de Bock GH, et al. Role of baseline nodule density and changes in density and nodule features in the discrimination between benign and malignant solid indeterminate pulmonary nodules. European Journal of Radiology 2009; 70:492-498
18. Silverman JD, Paul NS, Siewerdsen JH. Investigation of lung nodule detectability in low-dose 320-slice computed tomography. Med Phys 2009; 36:1700-1710
19. Paul NS, Siewerdsen JH, Patsios D, Chung TB. Investigating the low-dose limits of multidetector CT in lung nodule surveillance. Med Phys 2007; 34:3587-3595
20. Gavrielides MA, Kinnard LM, Myers KJ, Petrick N. Noncalcified lung nodules: volumetric assessment with thoracic CT. Radiology 2009; 251:26-37
21. Wang Y, van Klaveren RJ, van der Zaag-Loonen HJ, et al. Effect of nodule characteristics on variability of semiautomated volume measurements in pulmonary nodules detected in a lung cancer screening program. Radiology 2008; 248:625-631
22. Gietema HA, Wang Y, Xu D, et al. Pulmonary nodules detected at lung cancer screening: Interobserver variability of semiautomated volume measurements. Radiology 2006; 241:251-257
23. Das M, Ley-Zaporozhan J, Gietema HA, et al. Accuracy of automated volumetry of pulmonary nodules across different multislice CT scanners. European Radiology 2007; 17:1979-1984
24. Marten K, Dullin C, Machann W, et al. Comparison of flat-panel-detector-based CT and multidetector-row CT in automated volumetry of pulmonary nodules using an anthropomorphic chest phantom. British Journal of Radiology 2009; 82:716-723
25. Goodsitt MM, Chan HP, Way TW, Schipper MJ, Larson SC, Christodoulou EG. Quantitative CT of lung nodules: Dependence of calibration on patient body size, anatomic region, and calibration nodule size for single- and dual-energy techniques. Medical Physics 2009; 36:3107-3121
26. Bolte H, Riedel C, Jahnke T, et al. Reproducibility of computer-aided volumetry of artificial small pulmonary nodules in ex vivo porcine lungs. Investigative Radiology 2006; 41:28-35
27. Gietema HA, Schaefer-Prokop CM, Mali WPTM, Groenewegen G, Prokop M. Pulmonary nodules: Interscan variability of semiautomated volume measurements with multisection CT-influence of inspiration level, nodule size, and segmentation performance. Radiology 2007; 245:888-894
28. Wormanns D, Kohl G, Klotz E, et al. Volumetric measurements of pulmonary nodules at multi-row detector CT: in vivo reproducibility. European Radiology 2004; 14:86-92

29. Rampinelli C, De Fiori E, Raimondi S, Veronesi G, Bellomi M. In vivo repeatability of automated volume calculations of small pulmonary nodules with CT. American Journal of Roentgenology 2009; 192:1657-1661
30. Nietert PJ, Ravenel JG, Leue WM, et al. Imprecision in Automated Volume Measurements of Pulmonary Nodules and Its Effect on the Level of Uncertainty in Volume Doubling Time Estimation. Chest 2009; 135:1580-1587
31. Nasir K, Rubin J, Blaha MJ, et al. Interplay of coronary artery calcification and traditional risk factors for the prediction of all-cause mortality in asymptomatic individuals. Circulation-Cardiovascular Imaging 2012; 5:467-473
32. Oudkerk M, Stillman AE, Halliburton SS, et al. Coronary artery calcium screening: current status and recommendations from the European Society of Cardiac Radiology and North American Society for Cardiovascular Imaging. European radiology 2008; 18:2785-2807
33. Elias-Smale SE, Proenca RV, Koller MT, et al. Coronary calcium score improves classification of coronary heart disease risk in the elderly: the Rotterdam study. J Am Coll Cardiol 2010; 56:1407-1414
34. Greenland P, Bonow RO, Brundage BH, et al. ACCF/AHA 2007 clinical expert consensus document on coronary artery calcium scoring by computed tomography in global cardiovascular risk assessment and in evaluation of patients with chest pain: a report of the American College of Cardiology Foundation Clinical Expert Consensus Task Force (ACCF/AHA Writing Committee to Update the 2000 Expert Consensus Document on Electron Beam Computed Tomography). Circulation 2007; 115:402-426
35. Jain A, McClelland RL, Polak JF, et al. Cardiovascular Imaging for Assessing Cardiovascular Risk in Asymptomatic Men Versus Women The Multi-Ethnic Study of Atherosclerosis (MESA). Circulation-Cardiovascular Imaging 2011; 4:8-15
36. Berrington de Gonzalez A, Mahesh M, Kim KP, et al. Projected cancer risks from computed tomographic scans performed in the United States in 2007. Arch Intern Med 2009; 169:2071-2077
37. Oei HH, Vliegenthart R, Hofman A, Oudkerk M, Witteman JC. Risk factors for coronary calcification in older subjects. The Rotterdam Coronary Calcification Study. European heart journal 2004; 25:48-55
38. Mets OM, de Jong PA, Prokop M. Computed tomographic screening for lung cancer: an opportunity to evaluate other diseases. JAMA : the journal of the American Medical Association 2012; 308:1433-1434
39. Greuter MJW, Groen JM, Nicolai LJ, Dijkstra H, Oudkerk M. A model for quantitative correction of coronary calcium scores on multidetector, dual source, and electron beam computed tomography for influences of linear motion, calcification density, and temporal resolution: A cardiac phantom study. Medical Physics 2009; 36:5079-5088

40. White CS, Jerome S. Coronary calcium scoring on nongated chest CT: is it ready for prime time? Journal of cardiovascular computed tomography 2011; 5:119-121
41. Greuter MJW, Dorgelo J, Tukker WGJ, Oudkerk M. Study on motion artifacts in coronary arteries with an anthropomorphic moving heart phantom on an ECG-gated multidetector computed tomography unit. European Radiology 2005; 15:995-1007
42. Rabe KF, Hurd S, Anzueto A, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease - GOLD executive summary. American Journal of Respiratory and Critical Care Medicine 2007; 176:532-555
43. Patel BD, Coxson HO, Pillai SG, et al. Airway wall thickening and emphysema show independent familial aggregation in chronic obstructive pulmonary disease. American Journal of Respiratory and Critical Care Medicine 2008; 178:500-505
44. Nakano Y, Muro S, Sakai H, et al. Computed tomographic measurements of airway dimensions and emphysema in smokers correlation with lung function. American Journal of Respiratory and Critical Care Medicine 2000; 162:1102-1108
45. Coxson HO. Quantitative chest tomography in COPD research: chairman's summary. Proc Am Thorac Soc 2008; 5:874-877
46. Han MK, Kazerooni EA, Lynch DA, et al. Chronic obstructive pulmonary disease exacerbations in the COPDGene study: Associated radiologic phenotypes. Radiology 2011; 261:274-282
47. Matsuoka S, Yamashiro T, Washko GR, Kurihara Y, Nakajima Y, Hatabu H. Quantitative CT assessment of chronic obstructive pulmonary disease. Radiographics 2010; 30:55-66
48. Ginsburg SB, Lynch DA, Bowler RP, Schroeder JD. Automated texture-based quantification of centrilobular nodularity and centrilobular emphysema in chest CT images. Acad Radiol 2012; 19:1241-1251
49. Ceresa M, Bastarrika G, de Torres JP, et al. Robust, standardized quantification of pulmonary emphysema in low dose CT exams. Acad Radiol 2011; 18:1382-1390
50. Bartel ST, Bierhals AJ, Pilgram TK, et al. Equating quantitative emphysema measurements on different CT image reconstructions. Medical physics 2011; 38:4894-4902
51. Nakano Y, Sakai H, Muro S, et al. Comparison of low attenuation areas on computed tomographic scans between inner and outer segments of the lung in patients with chronic obstructive pulmonary disease: Incidence and contribution to lung function. Thorax 1999; 54:384-389
52. O'Donnell RA, Peebles C, Ward JA, et al. Relationship between peripheral airway dysfunction, airway obstruction, and neutrophilic inflammation in COPD. Thorax 2004; 59:837-842

53. Shaker SB, Maltbaek N, Brand P, Haeussermann S, Dirksen A. Quantitative computed tomography and aerosol morphometry in COPD and alpha-antitrypsin deficiency. European Respiratory Journal 2005; 25:23-30
54. Lamers RJ, Thelissen GR, Kessels AG, Wouters EF, van Engelshoven JM. Chronic obstructive pulmonary disease: evaluation with spirometrically controlled CT lung densitometry. Radiology 1994; 193:109-113
55. Zompatori M, Fasano L, Pacilli AM, Battista G, Cavina M, Pezzi S. Automatic evaluation of total lung capacity and of emphysema involvement with spiral computerized tomography (CT) in obstructive pneumonia. La Radiologia medica 2001; 101:18-24
56. Shaker SB, Dirksen A, Bach KS, Mortensen J. Imaging in chronic obstructive pulmonary disease. COPD-Journal of Chronic Obstructive Pulmonary Disease 2007; 4:143-161
57. Mets OM, de Jong PA, van Ginneken B, Gietema HA, Lammers JWJ. Quantitative computed tomography in COPD: possibilities and limitations. Lung 2012; 190:133-145
58. Weinheimer O, Achenbach T, Bletz C, Dueber C, Kauczor HU, Heussel CP. About objective 3-D analysis of airway geometry in computerized tomography. IEEE Transactions on Medical Imaging 2008; 27:64-74
59. de Jong PA, Muller NL, Pare PD, Coxson HO. Computed tomographic imaging of the airways: relationship to structure and function. Eur Resp J 2005; 26:140-152
60. Kosciuch J, Krenke R, Gorska K, Zukowska M, Maskey-Warzechowska M, Chazan R. Relationship between airway wall thickness assessed by high-resolution computed tomography and lung function in patients with asthma and chronic obstructive pulmonary disease. Journal of Physiology and Pharmacology 2009; 60 Suppl 5:71-76
61. Washko GR, Criner GJ, Mohsenifar Z, et al. Computed Tomographic-Based Quantification of Emphysema and Correlation to Pulmonary Function and Mechanics. COPD-Journal of Chronic Obstructive Pulmonary Disease 2008; 5:177-186
62. Akira M, Toyokawa K, Inoue Y, Arai T. Quantitative CT in chronic obstructive pulmonary disease: Inspiratory and expiratory assessment. American Journal of Roentgenology 2009; 192:267-272
63. Matsuoka S, Kurihara Y, Yagihashi K, Nakajima Y. Quantitative assessment of peripheral airway obstruction on paired expiratory/inspiratory thin-section computed tomography in chronic obstructive pulmonary disease with emphysema. Journal of Computer Assisted Tomography 2007; 31:384-389
64. Orlandi I, Moroni C, Camiciottoli G, et al. Spirometric-gated computed tomography quantitative evaluation of lung emphysema in chronic obstructive pulmonary disease - A comparison of 3 techniques. Journal of Computer Assisted Tomography 2004; 28:437-442

65. Mets OM, Buckens CF, Zanen P, et al. Identification of chronic obstructive pulmonary disease in lung cancer screening computed tomographic scans. JAMA : the journal of the American Medical Association 2011; 306:1775-1781
66. Pelkonen M. Smoking: relationship to chronic bronchitis, chronic obstructive pulmonary disease and mortality. Curr Opin Pulm Med 2008; 14:105-109
67. Guerra S, Sherrill DL, Venker C, Ceccato CM, Halonen M, Martinez FD. Chronic bronchitis before age 50 years predicts incident airflow limitation and mortality risk. Thorax 2009; 64:894-900
68. Huchon GJ, Vergnenegre A, Neukirch F, Brami G, Roche N, Preux PM. Chronic bronchitis among French adults: high prevalence and underdiagnosis. Eur Resp J 2002; 20:806-812
69. Hogg JC, Timens W. The pathology of chronic obstructive pulmonary disease. Annual review of pathology 2009; 4:435-459
70. Hogg JC. Pathophysiology of airflow limitation in chronic obstructive pulmonary disease. Lancet 2004; 364:709-721
71. Mullen JB, Wright JL, Wiggs BR, Pare PD, Hogg JC. Reassessment of inflammation of airways in chronic bronchitis. Br Med J (Clin Res Ed) 1985; 291:1235-1239
72. Hamada S, Ito Y, Imai S, Oguma T, Niimi A, Mishima M. Effect of inhaled corticosteroid therapy on CT scan-estimated airway dimensions in a patient with chronic bronchitis related to ulcerative colitis. Chest 2011; 139:930-932

Chapter 2

A Practical Approach to Radiological Evaluation of CT Lung Cancer Screening Examinations

Accepted by Cancer Imaging

Xueqian Xie[12] / Marjolein A. Heuvelmans[12] / Peter M.A. van Ooijen[12]
Matthijs Oudkerk[12] / Rozemarijn Vliegenthart[12]

University of Groningen, University Medical Center Groningen
[1]Department of Radiology, [2]Center for Medical Imaging – North East Netherlands

Abstract

Lung cancer is the most common cause of cancer-related death in the world. The Dutch-Belgian Randomized Lung Cancer Screening Trial (Dutch acronym: NELSON) was launched to investigate whether screening for lung cancer by low-dose multi-detector computed tomography (CT) in high risk subjects will lead to a decrease of lung cancer mortality. The NELSON lung nodule management is based on volumetry and volume-doubling time assessment. Evaluation of CT examinations in lung cancer screening can also include assessment of coronary calcification, emphysema and airway wall, markers for major diseases that share risk factors with lung cancer. In this review, a practical approach to the radiological evaluation of CT lung cancer screening examinations is described.

Introduction

Lung cancer is the primary cancer in males and the second in females, comprising 18% of the total number of deaths [1]. Despite advances in treatment, the 5-year survival rate is still only 15% or even less, as many lung cancers are found at a relatively late stage [2]. Low-dose computed tomography (CT) was proposed as a promising screening method for early detection of lung cancer.

The Dutch-Belgian Randomized Lung Cancer Screening Trial (Dutch acronym: NELSON) was launched in 2003. The hypothesis of the NELSON trial is that lung cancer screening by low-dose spiral CT will reduce 10-year lung cancer mortality by 25% in a high risk population. Details on participant recruitment and CT acquisition protocol were described elsewhere [3]. In short, four rounds of low-dose CT screening took place in the year 1, 2, 4, and 6 for heavy (ex-) smokers between 50 and 75 years of age. The NELSON lung nodule management is based on volumetry and volume-doubling time assessment. Thin-slice thoracic CT images were acquired with a slice thickness of 1 mm, and a slice interval of 0.7 mm, allowing for volume measurements of pulmonary nodules.

Evaluation of CT examinations in lung cancer screening can also include assessment of coronary calcification, emphysema and airway wall, markers for major diseases that share risk factors with lung cancer [4]. In this review, a practical approach to the radiological evaluation of CT lung cancer screening examinations is described, including evaluation of pulmonary nodules and nonnodular diseases.

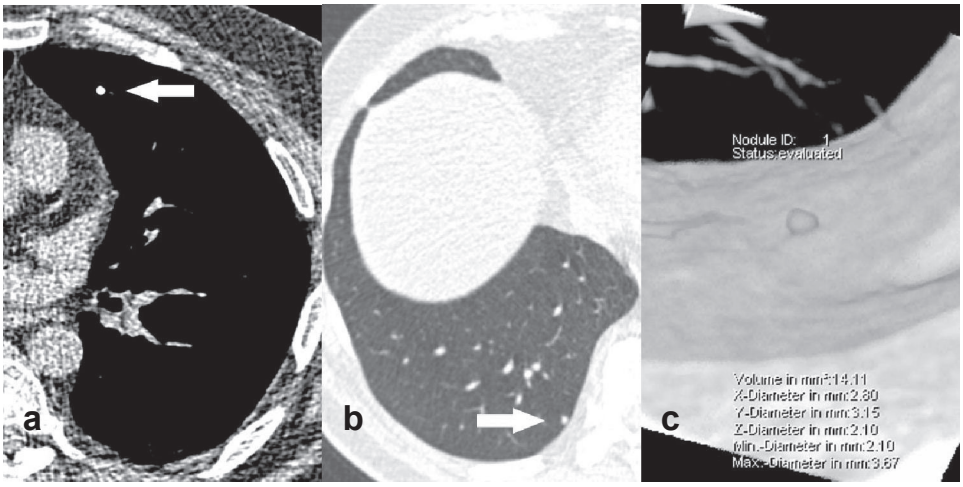


Figure 1 A complete calcified nodule is considered as benign (**a**). A very small (14 mm^3) nodule is not further evaluated (**b** and **c**).

Pulmonary nodule evaluation

Initial assessment

The assessment starts with evaluating whether a newly detected nodule has purely benign characteristics like benign calcifications or is very small (< 15 mm^3) (Figure 1). If so, the nodule is categorized as benign, and not further evaluated. If the nodule cannot directly be defined as benign, the nodule is further evaluated. Next, the density of the lung nodule is assessed. A nodule can be solid, partial-solid, or nonsolid.

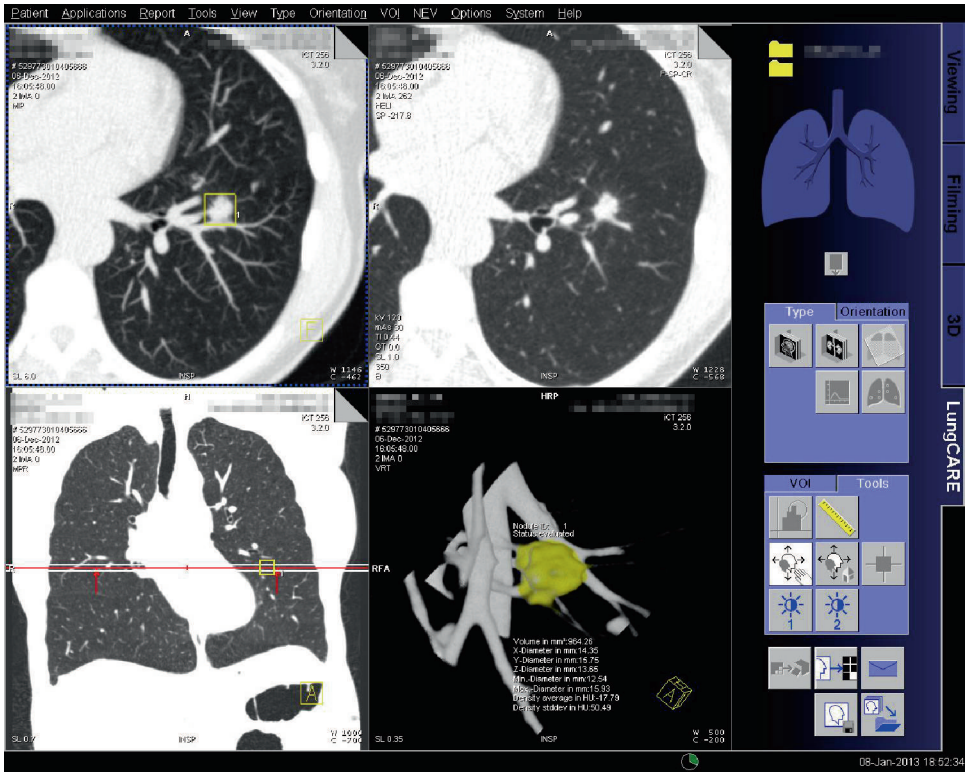


Figure 2 Screen capture of dedicated software (LungCare, Siemens, Forchheim, Germany) to semi-automatically measure the volume of a solid nodule.

Size-based evaluation

Evaluation of nodule size is essential to determine nodule growth. Evaluation methods for solid, partial-solid and nonsolid nodules are different.

Semi-automated volumetry measurements are utilized for segmentable nodules (Figure 2), e.g., solid nodules and the solid part of partial-solid nodules. In NELSON, approximately 98% of the total nodules have been found to be solid, thus are potential to be as-

sessed using semi-automated software [5]. In case of inappropriate segmentation, the reader is allowed to manually modify the segmentation for more accurate segmentation, which therefore overrules the automatically generated volume.

Manual measurement of diameters is performed for nonsegmentable nodules (Figure 3), e.g., nonsolid nodules and the nonsolid part of partial-solid nodules. Although partial-solid and nonsolid nodules constitute the minority of nodules that are detected, the frequency of malignancy is higher [6].

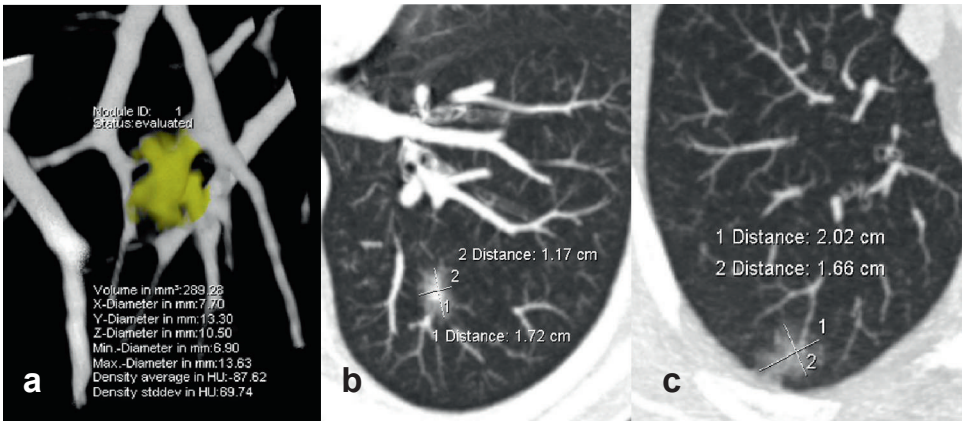


Figure 3 Semi-automated volumetry is performed for the solid part of a partial-solid nodule (**a**), and manual measurement of diameters is used for the nonsolid part of this partial-solid nodule (**b**). Manual measurement of diameters is performed for a nonsolid nodule (**c**).

In-vivo interscan variability for nodule size evaluation is inevitable. Based on validation studies with repeated low-dose CT on the same day, in which the measurement error was maximally 25%, nodule growth is defined as a change in volume of at least 25% between two subsequent examinations [7].

Additional nonsize based evaluation

Besides the nodule density, attachment type, shape, margin and location should be taken into account when evaluating a pulmonary nodule (Figure 4). Firstly, the attachment of nodules (peri-fissural, vessel-attached, pleural-based and intraparenchymal) is evaluated. Although peri-fissural nodules may show growth at follow-up CT, the malignancy potential of peri-fissural nodules is low [8]. Secondly, the shape of nodules (spherical and nonspherical) is evaluated. Nonspherical shape has been found to increase the likelihood of malignancy, rather than spherical shape [9]. Thirdly, the margin of nodules (smooth, lobulated, spiculated and others) is assessed. In a subgroup of NELSON with 469 solid intraparen-

chymal nodules, a lobulated or spiculated margin increased the likelihood for malignancy, rather than the smooth margin [10]. At last, nodule location is defined by the pulmonary segment and according to distance to pleura: peripheral nodules are defined as < 1/3 from total distance hilus-costal pleura, and central nodules are defined as > 2/3. In-between nodules are defined as between 1/3 to 2/3 from total distance hilus-costal pleura.

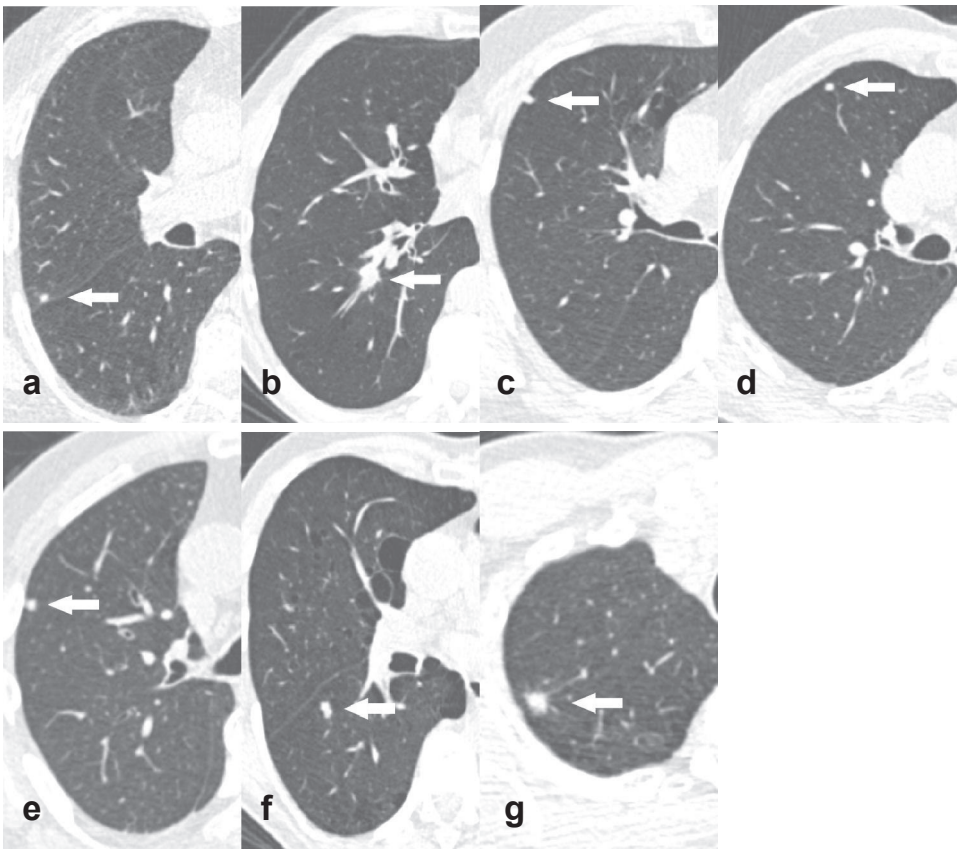


Figure 4 Example images of four nodule attachment types: peri-fissural (**a**), vessel-attached (**b**), pleural-based (**c**) and intraparenchymal (**d**). And of three margin types: smooth (**e**), lobulated (**f**) and spiculated (**g**).

Reading

A single CT lung cancer screening evaluation by an experienced reader seems sufficient. In the first to third round of the NELSON trial, images were evaluated twice. The second readers were unaware of the conclusion of the first reader. In case of discrepancy, a third reader made the final decision. However, based on the results from the first three rounds, no statistically significant benefit was found for consensus double reading for lung cancer

with the use of a nodule management strategy based on semi-automated volumetry measurements [11]. Thereafter, in the fourth round, only one reading was performed by one of the two radiologists with experiences of at least 8 years.

Images are interpreted on a workstation for pulmonary nodules and nonnodular diseases (in the NELSON trial: Leonardo, Siemens, Forchheim, Germany), both at lung window and mediastinal settings. When pulmonary nodules are identified, a dedicated software package (in the NELSON trial: LungCare, Siemens, Forchheim, Germany) is utilized for semi-automated nodule volumetry measurements. In case of nonsegmentable nodules, the reader should manually measure the lesion in diameter. In the LungCare software package, previous and current images are displayed simultaneously on the same screen for comparison. Besides the evaluation of nodule size, nodule characteristics (attachment type, margin, etc.) are then also evaluated.

Table 1 Nodule categorization based on size and density (new nodules) and growth rate (existing nodules) in the NELSON trial

Category	Definition		
NODCAT 1	A benign nodule (with fat/benign calcifications) or other benign abnormalities		
NODCAT 2	A nodule, smaller than NODCAT3, not belonging to NODCAT1		
	Solid	Partial solid	Nonsolid
NODCAT 3	$50 \leq V \leq 500$ mm^3 Pleural based: $5 \leq d_{min} \leq 10$ mm	Solid component: $50 \leq V \leq 500$ mm^3 Nonsolid component: $d_{mean} \geq 8$ mm	$d_{mean} \geq 8$ mm
NODCAT 4	$V > 500$ mm^3 Pleural based: $d_{min} > 10$ mm	Solid component: $V > 500$ mm^3	(nonexisting category)
GROWCAT A	VDT > 600 days		
GROWCAT B	400 ≤ VDT ≤ 600 days		
GROWCAT C	VDT < 400 days, or new solid component in nonsolid lesion		

V = volume; d_{min} = minimal diameter; d_{mean} = mean diameter; VDT = volume doubling time.

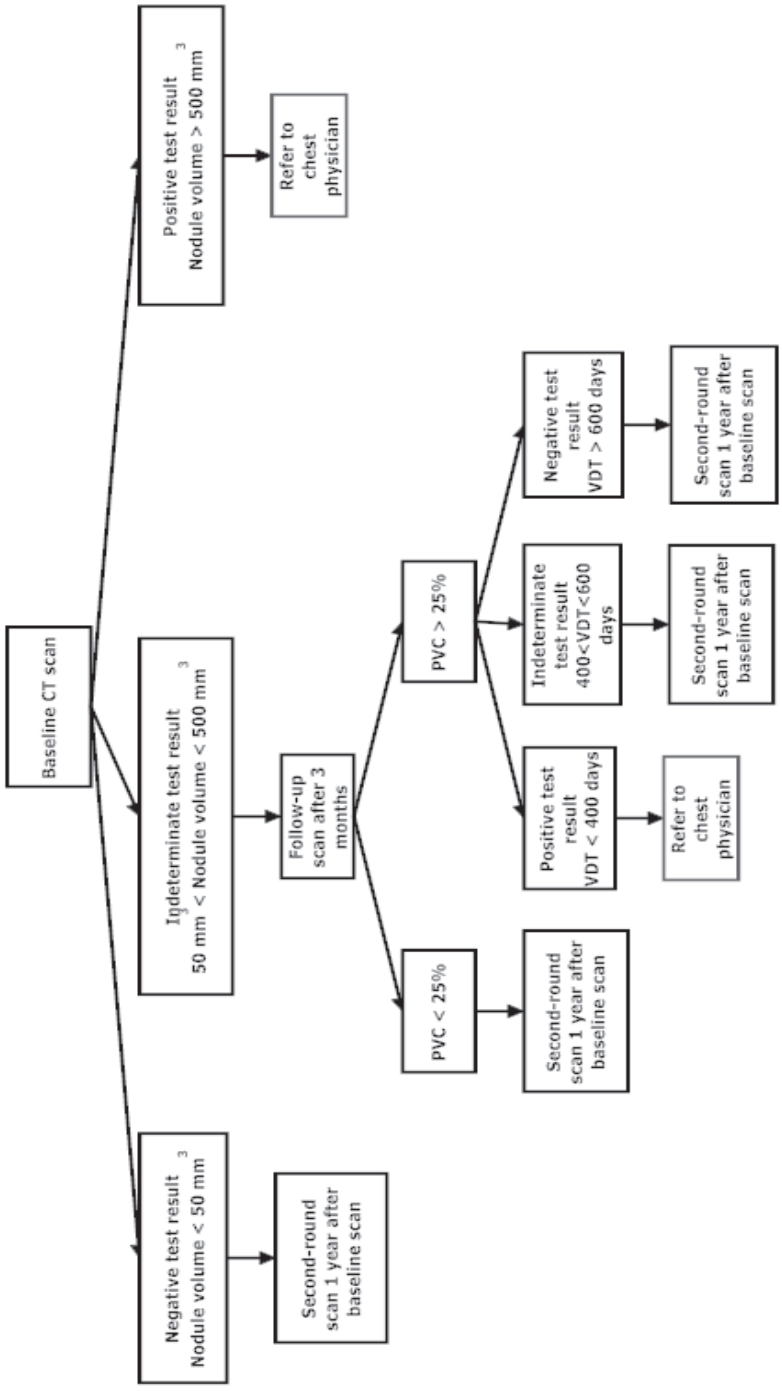


Figure 5 Decision tree of baseline examination in the NELSON trial.

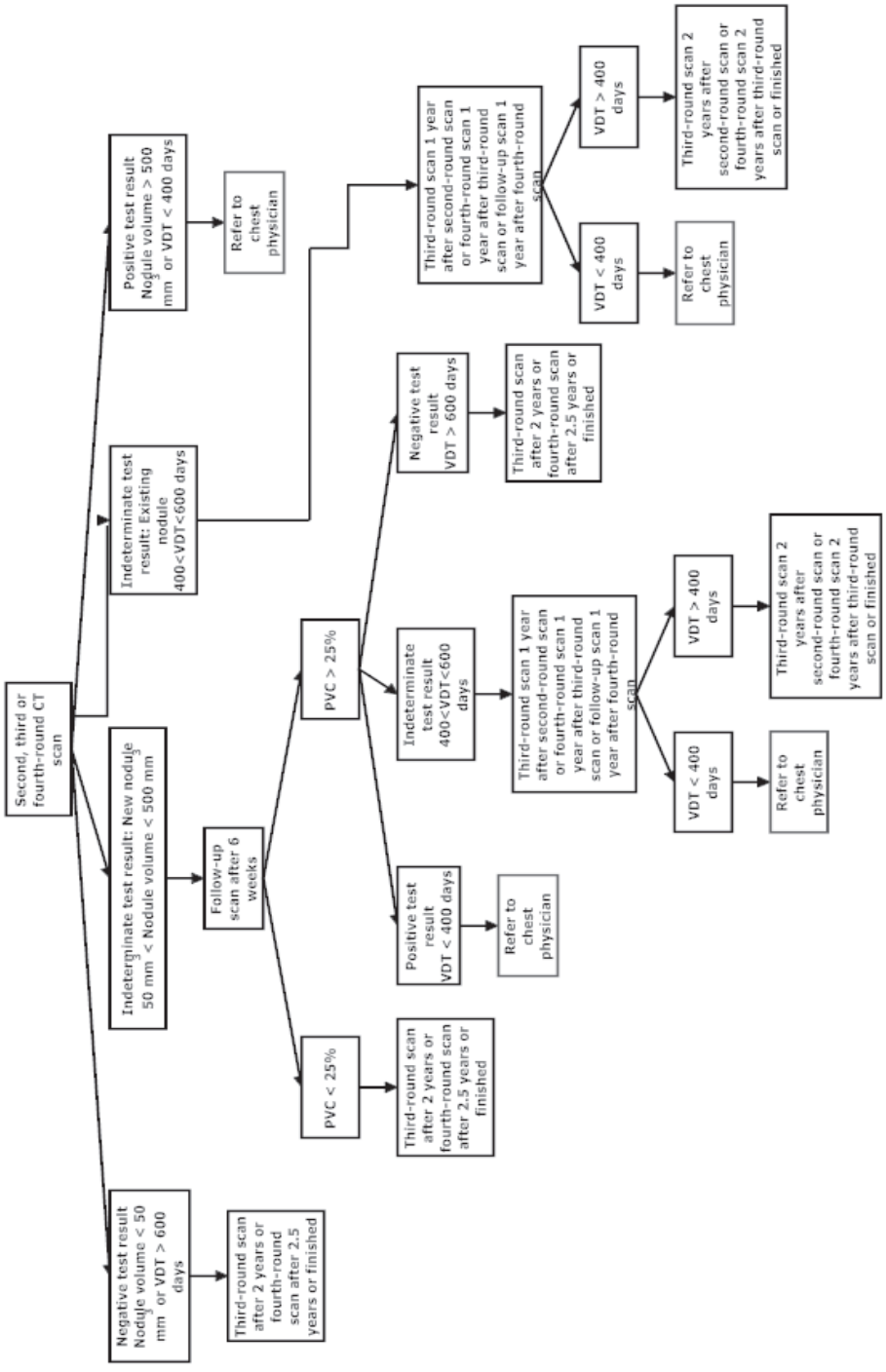


Figure 6 Decision tree of second, third and fourth round examination in the NELSON trial.

Lung nodule decision management

Newly detected lung nodules are divided into four categories (NODCAT 1 to 4), based on nodule density and size. In subsequent screening rounds, pre-existing nodules are defined as three categories (GROWCAT A to C), based on nodule growth in terms of volume-doubling time (Table 1). For newly detected nodules, the test result (negative, indeterminate and positive) is based on the highest NODCAT. For pre-existing nodules, the test result is based on highest GROWCAT. A negative result indicates that no further workup is needed. The participant is then invited to undergo the regular next-round CT. An indeterminate result requires a follow-up examination after 6 weeks (for incidence screening) to 3 months (for baseline screening). A positive result necessitates referral to a pulmonologist for work-up and diagnosis. The decision trees for baseline, second, third and fourth round are shown in Figures 5 and 6. An example of a growing nodule resulting in lung cancer is shown in Figure 7.

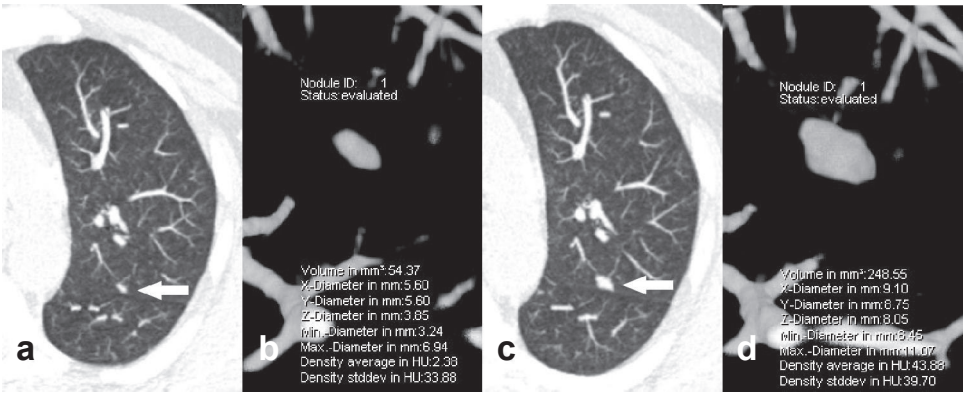


Figure 7 A new pulmonary nodule was found in apicoposterior segment of left superior lobe in the third round (**a** and **b**). The volume was 54 mm^3 (NODCAT 3). The nodule was nonspherical, lobulated and solid. On the fourth round, 3 years later (**c** and **d**), the nodule volume increased to 249 mm^3, and the volume-doubling time was 284 days (GROWCAT C). Thus, this participant was considered as a positive case and was referred to a pulmonologist. Finally this nodule was confirmed as lung cancer.

NELSON management system

In the NELSON trial, evaluation results are exported into the web-based NELSON management system (Figure 8). Nodule characteristics, volume and diameter are recorded. Nodule volume is automatically compared to the previous data to calculate the percentage volume change and the volume-doubling time in days. Finally, the system makes a suggestion for categorization of pulmonary nodules.

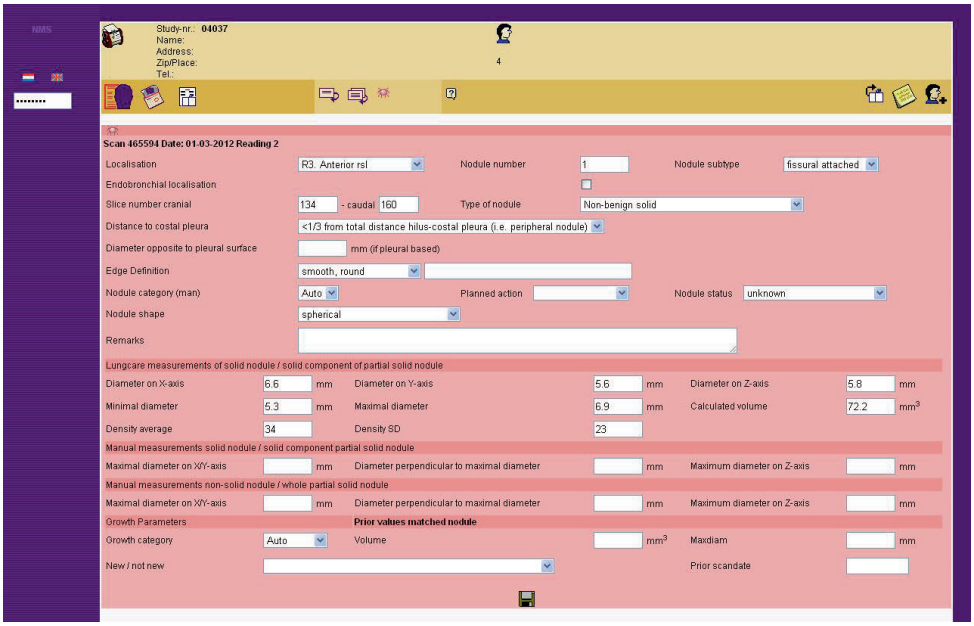


Figure 8 Screen capture of the web-based NELSON management system.

Table 2 Nonnodular radiological findings in lung cancer CT screening

	Not clinically significant	Clinically significant
Disease	Aortic calcium	Adrenal lesion
	Bronchiectasis	Aortic aneurysm
	Coronary artery calcifications	Bone destruction
	Emphysema	Liver lesions
	Pulmonary fibrosis	Mass (thyroid, breast, abdominal, etc.)
	Lymph node enlargement	Pleural fluid
	Pleural calcifications	Pneumonia
	Pleural plaques	Segmental or larger atelectasis

Nonnodular diseases

Besides lung cancer originating from pulmonary nodules, over 14% of participants in lung cancer screening have been found to have clinically significant diseases, such as cardiovascular diseases and pulmonary diseases [12]. Aging and smoking, the two major risk factors for lung cancer, are also the main contributors to the development and progression of

coronary calcification and emphysema [13,14]. It is important to review the CT screening examination for coronary artery calcification and emphysema. These important thoracic findings can be evaluated quantitatively, see the description below.

A list of nonnodular radiological findings that were initially reported in the NELSON trial is given in Table 2. Some severe diseases have been detected in the NELSON screen group, e.g., abdominal aortic aneurysm and renal cancer (Figure 9). However, the prevalence of clinically significant findings in the NELSON trial is only 1% and the benefit for systematically searching for these additional findings has been found to be neglectable [15].

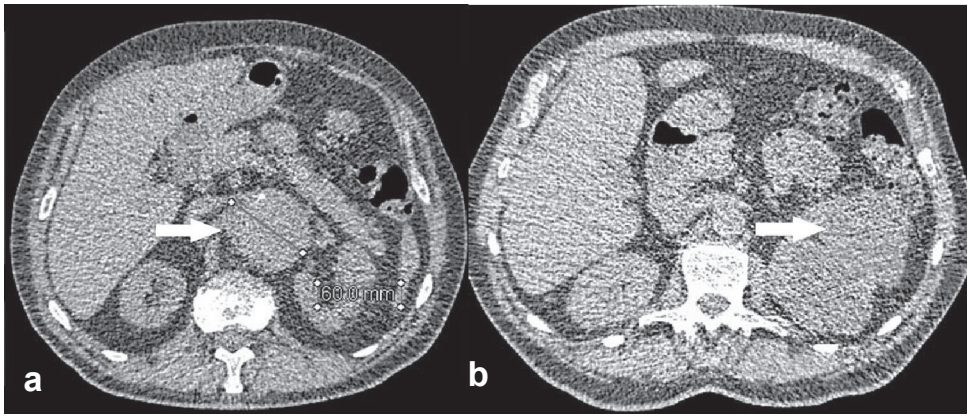


Figure 9 Incidentally found abdominal aortic aneurysm (**a**). The maximal diameter was measured as 6 cm. Incidentally found left renal mass (**b**), which was subsequently proved as renal carcinoma.

Coronary artery calcification

Coronary artery calcification is a frequent finding in the NELSON screening group, with a prevalence of over 70% [16]. Calcium scoring as part of low-dose CT lung cancer screening can be used as an independent predictor of cardiovascular death and events [16,17]. For the analysis of coronary calcification, the raw data should be reconstructed into 3 mm thickness to improve interscan reproducibility and make the settings more comparable to the dedicated coronary calcium examination [18,19]. Then, calcium scoring can be per-formed using the method developed by Agatston [20]. An example of evaluating calcium score is shown in Figure 10.

Figure 10 Screen capture of dedicated software (Aquarius iNtuition Calcium Score, TeraRecon, Foster City, US) to evaluate calcium score.

Emphysema and airway wall

Emphysema is also a frequent finding in lung cancer screening. The two primary causes of chronic obstructive pulmonary disease (COPD) are emphysema and airway remodelling [21]. In a meta-analysis, CT measured emphysema and airway wall thickness were significantly correlated with airflow obstruction in COPD [22]. CT examinations obtained for lung cancer screening can identify participants with COPD, with a sensitivity of 63% and a specificity of 88% [23]. Among the parameters that can be used to quantify emphysema, percentage of lung attenuation area under -950 HU (%LAA-950), mean lung density (MLD) and 15 percentile point of lung density (Perc 15) are the most commonly utilized. Among the parameters to quantify airway wall thick-ness, wall area percentage (WA%) and wall thickness (WT) are the most commonly utilized [22]. Dedicated software is needed to obtain quantitative emphysema and airway wall measures (Figures 11 and 12).

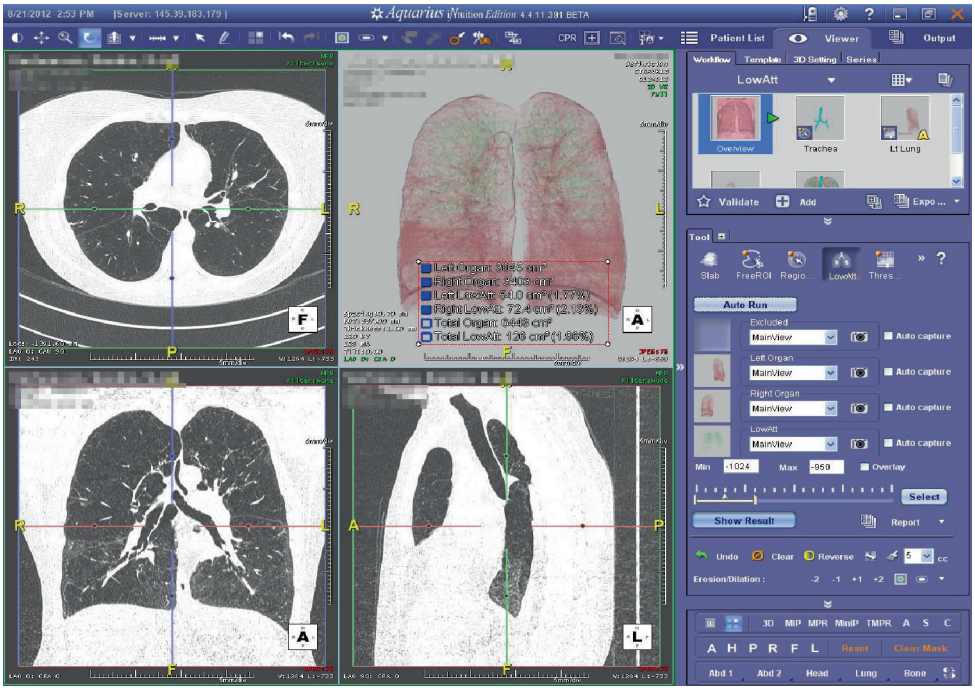


Figure 11 Screen capture of dedicated software to assess emphysema (Aquarius iNtuition LowAtt, TeraRecon, Foster City, US).

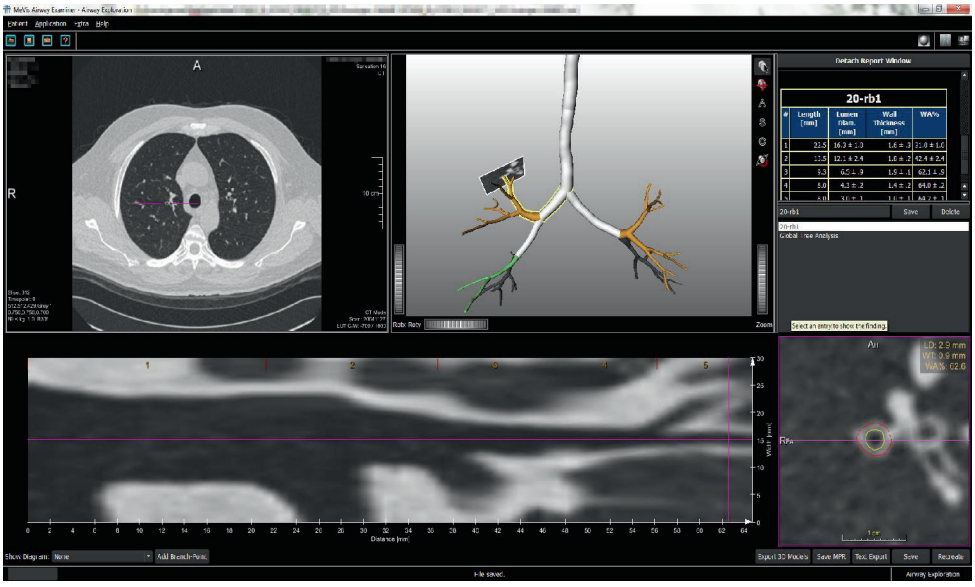


Figure 12 Screen capture of dedicated software to measure airway wall (Airway Examiner, Fraunhofer MEVIS, Bremen, Germany).

Conclusion

The NELSON trial is the first lung cancer screening trial in which nodules management is based on semi-automated volumetry measurements. The high resolution images acquired in low-dose thin-slice CT in NELSON are potential for evaluation of not only nodule volume, but also for interpretation and measurement of other thoracic structures. A 10-step practical approach to evaluate a participant was summarized in Table 3.

Cardiovascular diseases and COPD share risk factors, like aging and smoking, with lung cancer, and therefore, the incidence of these diseases in a lung cancer screening population is high. Besides the evaluation of lung nodules, low-dose thoracic CT is feasible to evaluate cardiac artery calcium, emphysema and airway wall thickness, and thereby can evaluate the risk for cardiovascular diseases and COPD. An integrated evaluation for these diseases is potential for participants in lung cancer screening.

Table 3 A 10-step approach to evaluate a participant in the NELSON trial

Step	Practical approach
1	Evaluate the presence of nodules (new or existing)
2	Exclude nodules with benign calcifications and nodules <15 mm^3
3	Determine the nodule density (solid, partial-solid, or nonsolid)
4	Measure the nodule (volume and diameters)
5	Determine the nodule characteristics (morphology, margin, location, lung segment)
6	In case of an existing nodule, compare to the previous examination, and determine volume doubling time
7	Categorize the nodule (for a new nodule based on size, for an existing nodule based on growth)
8	Repeat step 3-7 for each additional nodule
9	Review for coronary calcification, emphysema, and other findings
10	Derive test result (negative, indeterminate or positive)

References

1. Ferlay J, Shin HR, Bray F, et al. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer 2010; 127: 2893-2917.
2. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. CA Cancer J Clin 2012; 62: 10-29.
3. Ru Zhao Y, Xie X, de Koning HJ, et al. NELSON lung cancer screening study. Cancer imaging 2011; 11: S79-84.
4. Lozano R, Naghavi M, Foreman K, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2013; 380: 2095-2128.
5. van Klaveren RJ, Oudkerk M, Prokop M, et al. Management of lung nodules detected by volume CT scanning. New England Journal of Medicine 2009; 361: 2221-2229.
6. Henschke CI, Yankelevitz DF, Mirtcheva R, et al. CT screening for lung cancer: frequency and significance of part-solid and nonsolid nodules. AJR Am J Roentgenol 2002; 178: 1053-1057.
7. Gietema HA, Schaefer-Prokop CM, Mali WPTM, et al. Pulmonary nodules: interscan variability of semiautomated volume measurements with multisection CT-influence of inspiration level, nodule size, and segmentation performance. Radiology 2007; 245: 888-894.
8. Ahn MI, Gleeson TG, Chan IH, et al. Perifissural nodules seen at CT screening for lung cancer. Radiology 2010; 254: 949-956.
9. van't Westeinde SC, de Koning HJ, Xu DM, et al. How to deal with incidentally detected pulmonary nodules less than 10 mm in size on CT in a healthy person. Lung Cancer 2008; 60: 151-159.
10. Xu DM, van Klaveren RJ, de Bock GH, et al. Limited value of shape, margin and CT density in the discrimination between benign and malignant screen detected solid pulmonary nodules of the NELSON trial. European Journal of Radiology 2008; 68: 347-352.
11. Wang Y, van Klaveren RJ, de Bock GH, et al. No benefit for consensus double reading at baseline screening for lung cancer with the use of semiautomated volumetry software. Radiology 2012; 262: 320-326.
12. Jacobs PC, Mali WP, Grobbee DE, et al. Prevalence of incidental findings in computed tomographic screening of the chest: a systematic review. J Comput Assist Tomogr 2008; 32: 214-221.
13. McEvoy JW, Blaha MJ, Rivera JJ, et al. Mortality rates in smokers and nonsmokers in the presence or absence of coronary artery calcification. JACC-Cardiovascular Imaging 2012; 5: 1037-1045.

14. Maritz GS, Mutemwa M. Tobacco smoking: patterns, health consequences for adults, and the long-term health of the offspring. Global journal of health science 2012; 4: 62-75.
15. van de Wiel JCM, Wang Y, Xu DM, et al. Neglectable benefit of searching for incidental findings in the Dutch-Belgian lung cancer screening trial (NELSON) using low-dose multi-detector CT. Eur Radiol 2007; 17: 1474-1482.
16. Jacobs PC, Gondrie MJA, van der Graaf Y, et al. Coronary artery calcium can predict all-cause mortality and cardiovascular events on low-dose CT screening for lung cancer. AJR American Journal of Roentgenology 2012; 198: 505-511.
17. Jacobs PC, Gondrie MJ, Mali WP, et al. Unrequested information from routine diagnostic chest CT predicts future cardiovascular events. Eur Radiol 2011; 21: 1577-1585.
18. Jacobs PC, Isgum I, Gondrie MJ, et al. Coronary artery calcification scoring in low-dose ungated CT screening for lung cancer: interscan agreement. AJR Am J Roentgenol 2010; 194: 1244-1249.
19. Isgum I, Prokop M, Jacobs PC, et al. Automatic coronary calcium scoring in low-dose non-ECG-synchronized thoracic CT scans. Proceedings of the SPIE - The International Society for Optical Engineering 2010; 7624: 76240M.
20. Agatston AS, Janowitz WR, Hildner FJ, et al. Quantification of coronary artery calcium using ultrafast computed tomography. J Am Coll Cardiol 1990; 15: 827-832.
21. Huertas A, Palange P. COPD: a multifactorial systemic disease. Therapeutic Advances in Respiratory Disease 2011; 5: 217-224.
22. Xie XQ, de Jong PA, Oudkerk M, et al. Morphological measurements in computed tomography correlate with airflow obstruction in chronic obstructive pulmonary disease: systematic review and meta-analysis. Eur Radiol 2012; 22: 2085-2093.
23. Mets OM, Buckens CF, Zanen P, et al. Identification of chronic obstructive pulmonary disease in lung cancer screening computed tomographic scans. JAMA 2011; 306: 1775-1781.

Chapter 3

Sensitivity and Accuracy of Volumetry of Pulmonary Nodules on Low-Dose 16- and 64-Row Multi-Detector CT: an Anthropomorphic Phantom Study

European Radiology. 2013; 23: 139-147

Xueqian Xie[12] / Yingru Zhao[12] / Roland A. Snijder[1] / Peter M.A. van Ooijen[12]
Pim A. de Jong[3] / Matthijs Oudkerk[2] / Geertruida H. de Bock[4]
Rozemarijn Vliegenthart[12] / Marcel J.W. Greuter[1]

University of Groningen, University Medical Center Groningen, [1]Department of Radiology,
[2]Center for Medical Imaging – North East Netherland, [4]Department of Epidemiology;
[3]Department of Radiology, University Medical Center Utrecht

Abstract

Objective

To assess the sensitivity of detection and accuracy of volumetry by manual and semiautomated quantification of artificial pulmonary nodules in an anthropomorphic thoracic phantom on low-dose CT.

Materials and Methods

Fifteen artificial spherical nodules (diameter 3, 5, 8, 10 and 12 mm; CT densities -800, -630 and +100 HU) were randomly placed inside an anthropomorphic thoracic phantom. The phantom was examined on 16- and 64-row multidetector CT with a low-dose protocol. Two independent blinded observers screened for pulmonary nodules. Nodule diameter was measured manually, and volume calculated. For solid nodules (+100 HU), diameter and volume were also evaluated by semi-automated software. Differences in observed volumes between the manual and semi-automated method were evaluated by a t-test.

Results

Sensitivity was 100% for all nodules of > 5 mm and larger, 60 to 80% for solid and 0 to 20% for nonsolid 3-mm nodules. No false-positive nodules but high inter-observer reliability and inter-technique correlation were found. Volume was underestimated manually by 24.1 ± 14.0% for nodules of any density, and 26.4 ± 15.5% for solid nodules, compared with 7.6 ± 8.5% ($p < 0.01$) semi-automatically.

Conclusions

In an anthropomorphic phantom study, the sensitivity of detection is 100% for nodules of > 5 mm in diameter. Semi-automated volumetry yielded more accurate nodule volumes than manual measurements.

Introduction

The most common cause of cancer-related death is lung cancer. In 2000, lung cancer accounted for 17% of total cancer mortality [1]. Despite advances in treatment the 5-year survival rate is still only 15% or even less, as many lung cancers are found at a relatively late stage [2]. Potentially, screening for lung cancer may improve the prognosis. About a decade ago, low-dose computed tomography (CT) was proposed as a promising method to screen for lung cancer. Several cohort studies and randomised clinical trials using low-dose CT were started, aiming to investigate the effect of lung cancer screening on the distribution of tumour stages and eventually the effect on survival [3-5]. Recently, the initial, encouraging results of one of the largest trial cohorts were published: lung cancer screening by low-dose CT reduced lung cancer-specific mortality by 20% compared with chest radiography [3].

Efficient lung cancer screening depends on the accurate distinction between benign and malignant pulmonary lesions, but starts with sensitive observer detection of pulmonary nodules. There are only scarce validation data on the detectability of small pulmonary nodules by low-dose CT [6,7]. In one previous study using older generation CT equipment with slightly thicker collimation, the sensitivity of the detection of pulmonary nodules was evaluated in a thoracic phantom without pulmonary vessels, in which artificial nodules were placed at known locations. Pulmonary nodules as small as 2.4 mm in diameter could be detected by conventional dose CT [8]. However, no data exist about the observer sensitivity on current thin-slice low-dose CT for pulmonary nodules in an anthropomorphic pulmonary background with pulmonary vessels. In addition, limited data are available on the accuracy of volumetry of pulmonary nodules on low-dose CT. Therefore, the aim of this manuscript was to assess the sensitivity of detection of artificial pulmonary nodules on low-dose CT, randomly placed in an anthropomorphic pulmonary background with pulmonary vessels, and to determine the accuracy of volumetry of detected nodules by manual and semi-automated measurements.

Materials and Methods

An anthropomorphic thoracic phantom (Lungman, Kyoto Kagaku, Tokyo, Japan) with artificial thoracic wall, heart, mediastinum, diaphragm and lung with pulmonary vessels was used (Figure 1). The phantom consisted of an accurate life-size anatomical model of a male thorax with soft tissue substitute materials made of polyurethane resin composites and synthetic bones made of epoxy resin with X-ray absorption rates very close to those of human tissue. The space between the pulmonary vessels in the thoracic cavity consisted of air. In addition, we used 15 artificial spherical pulmonary nodules with a smooth surface in five

diameters (3, 5, 8, 10 and 12 mm, corresponding to a volume of 14, 65, 268, 523 and 904 mm^3) and 3 CT densities (-800, -630 and +100 Hounsfield Units [HU]). The artificial solid nodules were made of polyurethane resin and the artificial nonsolid nodules were made of polyurethane foam resin.

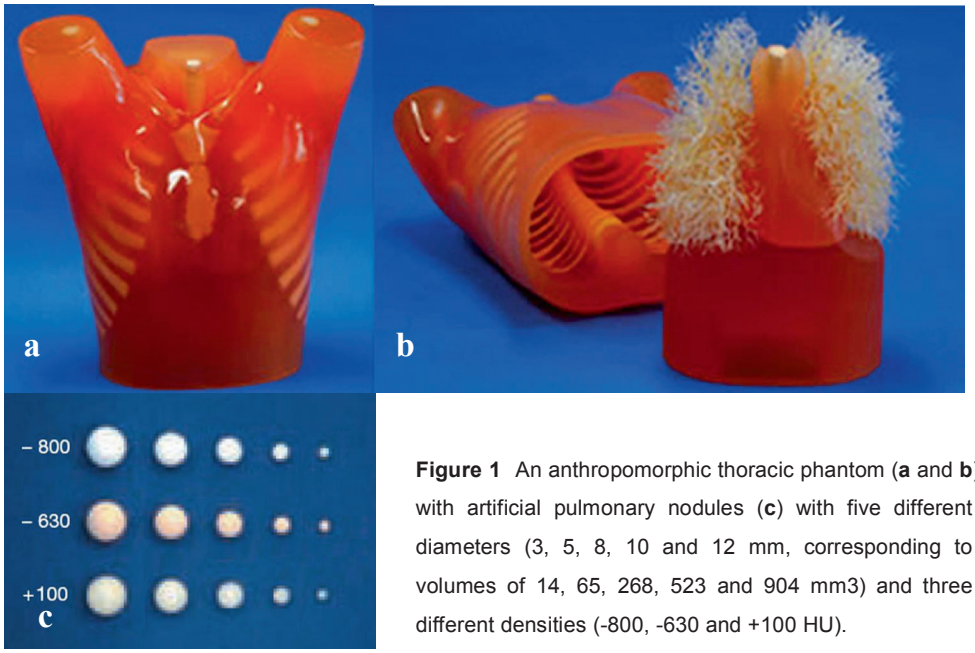


Figure 1 An anthropomorphic thoracic phantom (**a** and **b**) with artificial pulmonary nodules (**c**) with five different diameters (3, 5, 8, 10 and 12 mm, corresponding to volumes of 14, 65, 268, 523 and 904 mm3) and three different densities (-800, -630 and +100 HU).

16-row multi-detector CT (16-MDCT) and 64-row multi-detector CT (64-MDCT) (Sensation 16 and Sensation 64, respectively, Siemens, Forchheim, Germany) were used. A clinically used low-dose protocol for lung cancer screening was applied for image acquisition [9]: spiral acquisition at 120 kV, 20 mAs, 0.5 s rotation time, pitch 1.5 and collimation 16 × 0.75 mm and 2 × 32 × 0.6 mm. The CT images were reconstructed with a slice thickness of 1 mm and increment of 0.7 mm using a medium B30f kernel and a field of view 300 mm.

The artificial nodules were randomly positioned in both artificial lungs. All nodules were attached to pulmonary vessels. None of the nodules were attached to pleura or positioned sub-pleurally. A random, pre-determined set of 6 nodules was positioned in the artificial lungs, after which CT examination was performed. Each of the 15 nodules was examined in total 5 times, thus 13 different sets of nodules were positioned in the phantom (the last set consisted only of 3 nodules). The CT examination was, for each new phantom nodule set-up, repeated 5 times, with in between each examination a small translocation and rotation of the phantom to simulate participant movement. The thoracic phantom was

also examined five times without pulmonary nodules to serve as a control. Thus, per CT technique 70 examinations were performed, including 65 test examinations and 5 control examinations. The thoracic phantom was examined with the same settings for the two CT techniques. Furthermore, all the nodules were examined in air, on the CT table, to confirm the visibility on low-dose CT.

The reconstructed data were evaluated on a dedicated workstation (Leonardo, Siemens, Forchheim, Germany) by two independent observers, both with experience in thoracic diagnostic imaging for more than 8 years, who were blinded to information about the presence, properties and location of the artificial pulmonary nodules. All the examinations were read by both observers. The observers were instructed to review the images for the presence of nodules within clinically relevant time duration (about 2 minutes per examination). The observers were asked to report whether there were one or more pulmonary nodules present or not. If a potential nodule was observed, the images were compared with the images of the control CT examination without nodules to confirm it was not a false-positive finding caused by pulmonary background structures. Subsequently, the slice with maximal cross-sectional area of the nodule was selected. Then, the diameter and CT density were manually measured. The diameter was measured according to the Response Evaluation Criteria in Solid Tumors (RECIST) [10]. A region of interest was drawn as large as possible within the nodule border to measure the CT density. Because all nodules were spherical, the volume of the nodules could be easily calculated from the measured diameter. Additionally, a dedicated semi-automated software tool (LungCARE, Siemens, Forchheim, Germany) was used to measure the volume of the detected solid nodules (CT density +100 HU). The diameter and volume of identified nodules were automatically calculated by this three-dimensional volumetric assessment software.

Statistics

The sensitivity of detection of artificial pulmonary nodules was calculated for three densities (-800, -630 and +100 HU) and for 16- and 64-row multi-detector CT. The inter-observer reliability for both manual and semi-automated measurements was assessed using an intraclass correlation coefficient. The correlation of measurements between 16- and 64-row multi-detector CT was expressed as a Pearson's correlation coefficient. The inter-observer and inter-machine agreement of nodule volumetry was analysed using Bland-Altman plots. If there was a difference in the measurements between 16- and 64-row multi-detector CT, or between the two observers, an independent sample *t*-test was used. If there was a difference in measuring diameter and volume using the manual and semi-automated methods, a paired-samples *t*-test was used. To find the difference between the observed value (diameter, volume and density) and the actual value, a one-sample *t*-test was used. Results were given as mean ± standard deviation (SD). A $p < 0.05$ was considered to be statistically sig-

nificant. All statistical analyses were performed with a software package (SPSS 18.0.3, IBM, New York, USA).

Results

Representative CT images of the anthropomorphic thoracic phantom are shown in Figure 2. Nodules sized 5 mm in diameter and larger of all CT densities were detected by both observers on all examinations obtained with 16- and 64-MDCT. For nodules sized 3 mm in diameter, solid nodules (CT density +100 HU) were detected on 15 out of 25 examinations (60%) for both reviewers in 16-MDCT, and 15 (60%) and 20 (80%) examinations for the first and the second reviewer respectively in 64-MDCT. However, nonsolid nodules with CT density of -630 HU were only detected on 5 examinations out of 25 (20%) by the second reviewer on 16-MDCT. Nonsolid nodules with CT density of -800 HU were not detected on any examination (Figure 3). Each observed nodule was compared with the control examination to confirm the presence of the nodule; no false-positive nodules were found. All the nodules found were measured successfully using the manual or semi-automated method, except for one nodule with a diameter of 3 mm and a density of +100 HU for which segmentation by the semi-automated method failed. Furthermore, on the CT examinations of the nodules in air (without the phantom), all 15 nodules were visualised on low-dose CT.

The inter-observer reliability was very good with an intraclass correlation coefficient of 0.985 ($p < 0.001$) and 1.000 ($p < 0.001$) for manual and semi-automated measurement. The correlation of nodule measurements between 16-MDCT and 64-MDCT was high with a Pearson's correlation coefficient of 0.983 ($p < 0.001$) and 0.999 ($p < 0.001$) for manual and semi-automated measurements, respectively. An increasing relative inter-observer and inter-machine volumetry difference at smaller nodule size was found (Figures 4 and 5). The mean absolute value of relative inter-observer difference was 11.7 ± 14.4% and 3.3 ± 6.6% for manual and semi-automated volumetry, respectively. The mean absolute value of relative inter-scanner difference was 12.9 ± 12.4% and 4.0 ± 5.6%, respectively. There was no significant difference in the diameter measurements or CT density between the two techniques or between the two observers ($p > 0.05$).

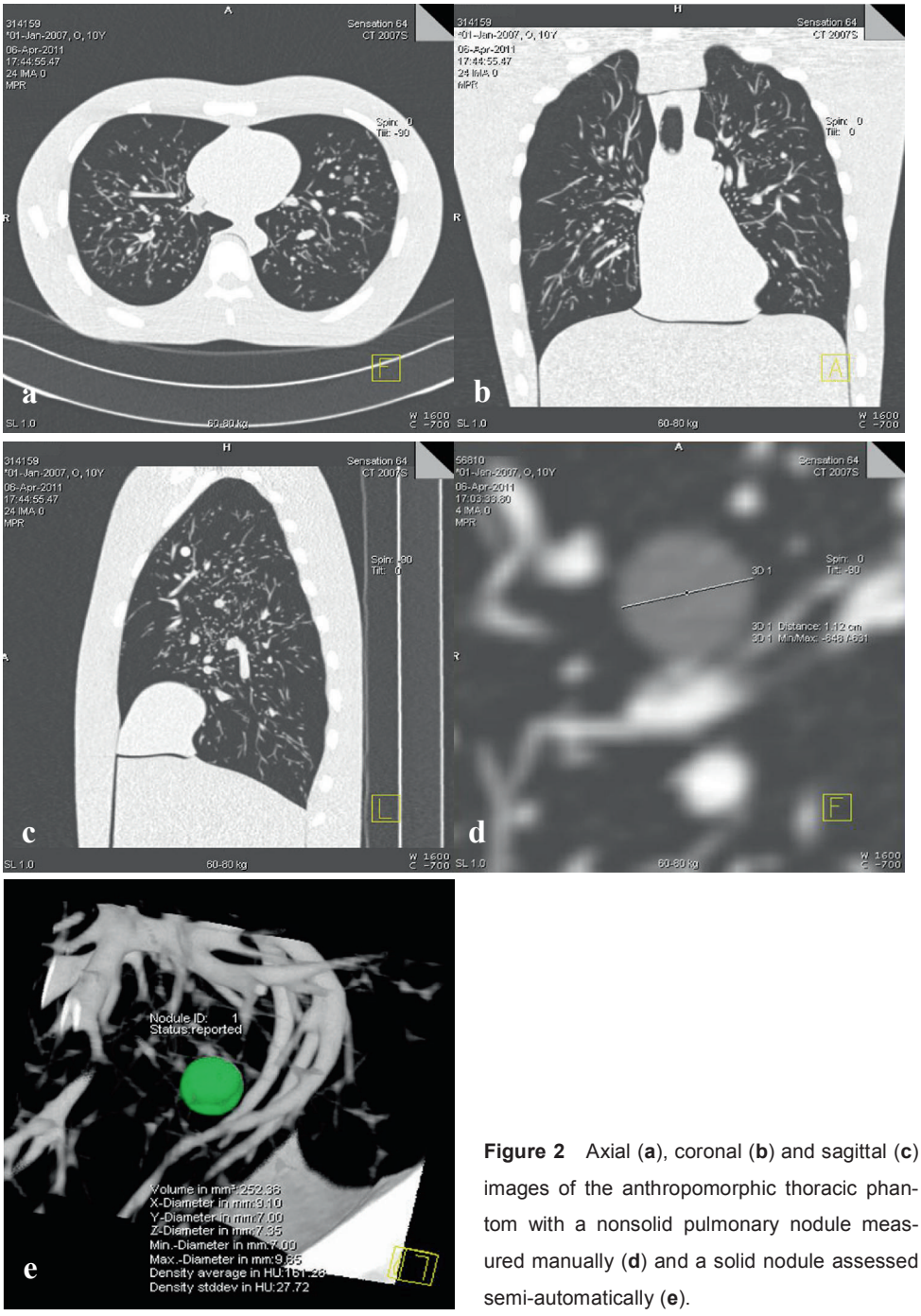


Figure 2 Axial (**a**), coronal (**b**) and sagittal (**c**) images of the anthropomorphic thoracic phantom with a nonsolid pulmonary nodule measured manually (**d**) and a solid nodule assessed semi-automatically (**e**).

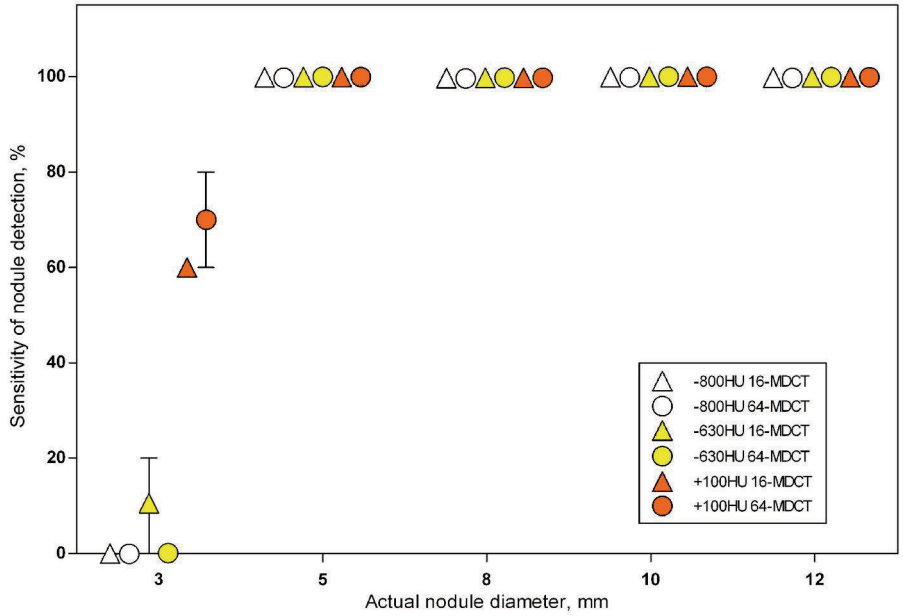


Figure 3 Sensitivity of the detection of artificial pulmonary nodules for three densities (-800, -630 and +100 HU) and two CT techniques (16-MDCT and 64-MDCT).

In both the manual and the semi-automated method, nodule diameter and volume were significantly underestimated compared with the actual properties ($p < 0.01$) (Table 1). An increasing underestimation of nodule diameter and volume at smaller nodule size was found (Figures 6 and 7). In diameter evaluation, the overall underestimation for nodules of any density was 9.2 ± 6.0% using manual method. For solid nodules, the underestimation was 10.1 ± 6.9% using the manual method, compared with 3.7 ± 7.1% ($p < 0.01$) using the semi-automated method. In volumetry, the overall underestimation for nodules of any density was 24.1 ± 14.0% using the manual method. For solid nodules, the underestimation was 26.4 ± 15.5% using the manual method, compared with 7.6 ± 8.5% ($p < 0.01$) using the semi-automated method.

The mean measured CT density was -813 ± 23 HU, -647 ± 9 HU and 123 ± 61 HU for nodule density of -800 HU, -630 HU and +100 HU respectively (Table 1), thus deviating 1.7 ± 2.3%, -2.7 ± 1.5% and 26 ± 57% from the expected density ($p < 0.01$).

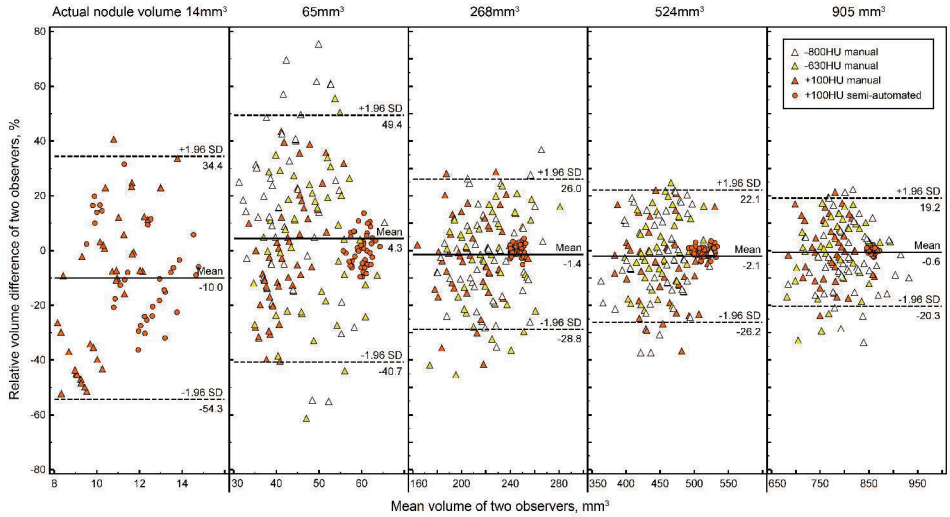


Figure 4 Bland-Altman plots for inter-observer agreement of the measured volume.

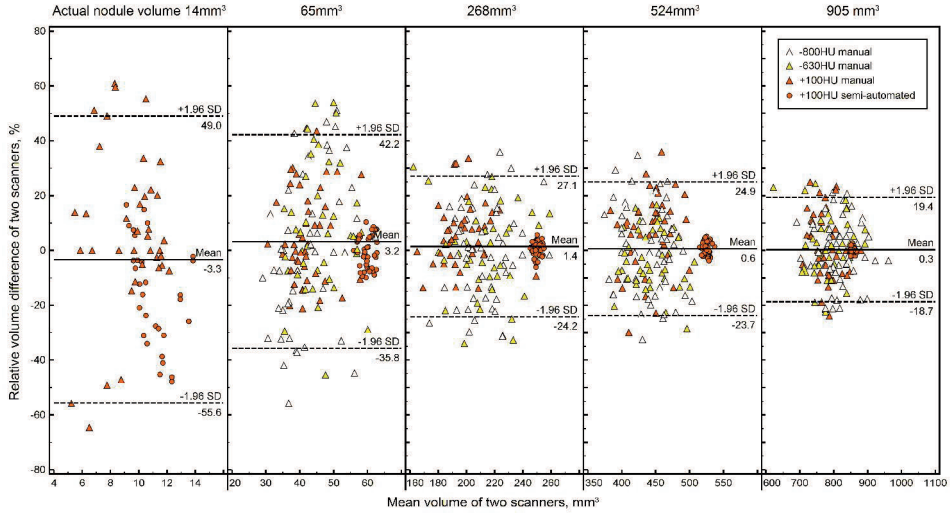


Figure 5 Bland-Altman plots for inter-scanner agreement of the measured volume.

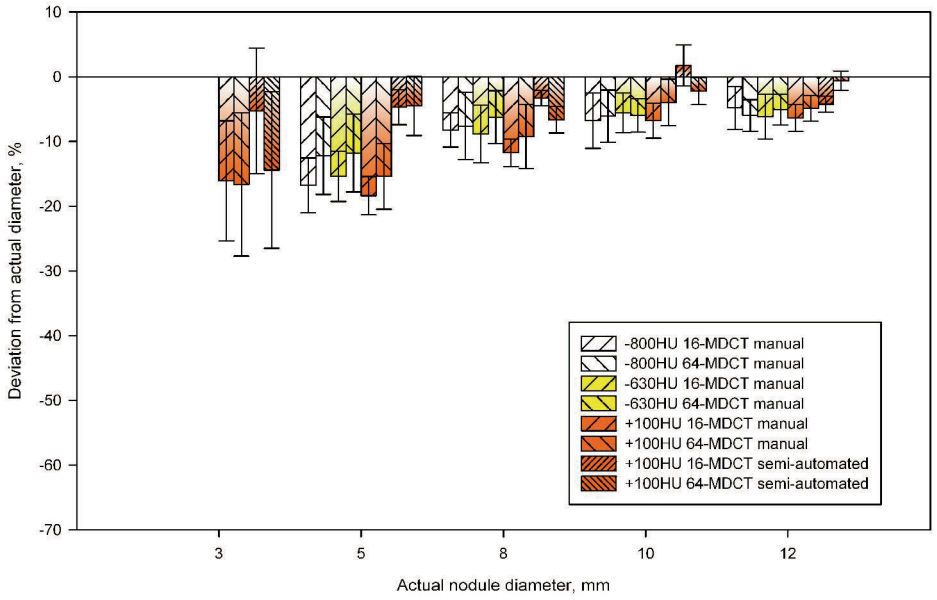


Figure 6 Deviation of the measured diameter from the actual diameter by manual measurement for nodules of -800, -630 and +100 HU, and by semi-automated measurement for nodules of +100 HU.

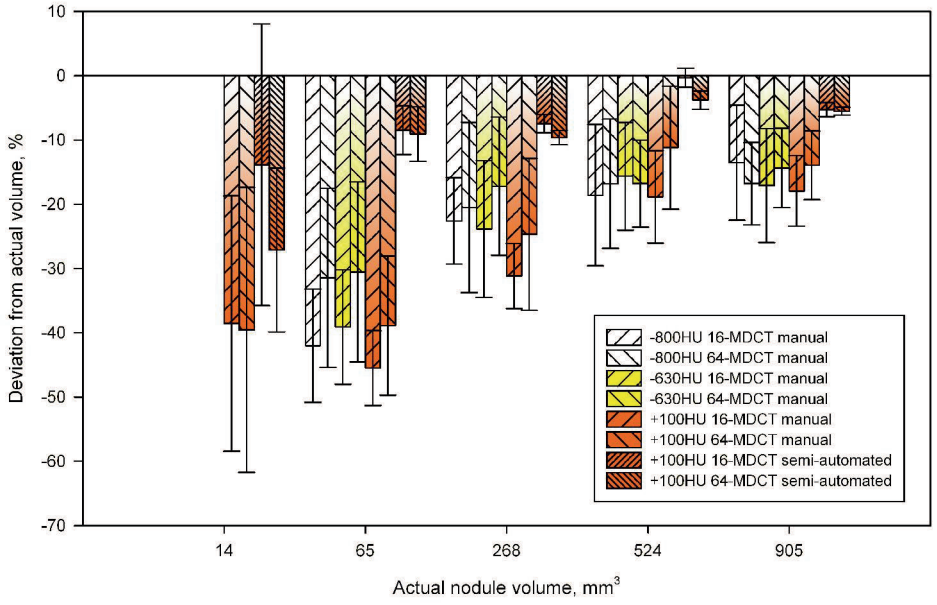


Figure 7 Deviation of the measured volume from the actual volume by manual volumetry for nodules of -800, -630 and +100 HU, in addition to by semi-automated volumetry for nodules of +100 HU.

Table 1 Measurements of diameter and volume of artificial pulmonary nodules by the manual and semi-automated methods

Actual property			Manual measurement			Semi-automated measurement	
Density, HU	Diameter, mm	Volume, mm^3	Density, HU	Diameter, mm	Volume, mm^3	Diameter, mm	Volume, mm^3
	3	14	n/d	n/d	n/d	n/a	n/a
	5	65	-811±15*	4.3±0.3*	41.4±8.2*	n/a	n/a
-800	8	268	-818±6*	7.4±0.3*	210±27*	n/a	n/a
	10	523	-819±18*	9.4±0.4*	431±54*	n/a	n/a
	12	904	-806±8*	11.4±0.3*	768±70*	n/a	n/a
	3	14	-639±25*	2.6±0.1*	9.3±1.3*	n/a	n/a
	5	65	-646±14*	4.3±0.3*	42.7±8.0*	n/a	n/a
-630	8	268	-646±5*	7.4±0.3*	213±29*	n/a	n/a
	10	523	-648±6*	9.4±0.3*	439±39*	n/a	n/a
	12	904	-650±6*	11.3±0.4*	763±69*	n/a	n/a
	3	14	-2±56*	2.4±0.3*	7.1±3.9*	2.6±0.4*†	10.9±2.4*†
	5	65	161±24*	4.2±0.2*	37.9±6.0*	4.8±0.2*†	59.7±2.6*†
+100	8	268	158±10*	5.9±1.2*	193±25*	7.6±0.2*†	245±4.5*†
	10	523	137±8*	9.5±0.3*	445±48*	9.8±0.3*†	513±12*†
	12	904	128±6*	11.3±0.3*	761±51*	11.7±0.3*†	856±8*†

*: $P<0.01$ when compared with actual properties.

†: $P<0.01$ when compared with manual measurement.

The measured values were averaged over 16-slice and 64-slice CT for both observers. n/a = not available; n/d = not detectable.

Discussion

In one of the first pulmonary nodule validation studies using an anthropomorphic thoracic phantom and current low-dose CT technology, we have shown that a clinically used lung cancer screening protocol with low-dose CT has 100% sensitivity of detection for spherical pulmonary nodules sized 5 mm in diameter and larger. In addition, we have shown that low-dose CT yields more accurate nodule volume measurements when using a semi-automated method than in the case of the manual method, with negligible underestimation of actual size, especially for small nodules.

We found a sensitivity of 100% for nodules with a diameter equal to or larger than 5 mm for all three nodule densities in this anthropomorphic thoracic phantom, and 60 to 80% and 0 to 20% for solid and nonsolid nodules with a diameter of 3 mm, respectively. This anthropomorphic thoracic phantom was also used in another study in which sensitivity of 95% for solid nodules and 74 – 81% for nonsolid nodules were found for nodule diameter equal to or larger than 5 mm [11]. Unlike the low-dose acquisition protocol for lung cancer screening in this study, the authors did not use a low-dose protocol, which limits comparability. In addition, the authors used 5-mm reconstructed slice thickness, compared with 1 mm in this study. It is well known that a larger slice width results in lower sensitivity of pulmonary nodules [6]. This anthropomorphic phantom was also used for an image database of nodules of diameter larger than 5 mm of several shapes, but results for nodule detectability and comparing between manual and semi-automated measurements were not reported [12]. In some nodule detectability studies, solid nodules with a diameter of 2 to 3 mm were detected in all cases [8,13,14]. In these studies, the nodules were placed in known order and examined in a thoracic phantom without lung vessels. On the other hand, in our study the artificial nodules were randomly positioned inside the lungs of an anthropomorphic thoracic phantom, thus limiting detection bias and strengthening the findings. As adjacent tissue can interfere with the nodule image reconstruction and reading, especially for nonsolid nodules, and because this interference increases with decreasing radiation dose, we expect that this interference explains why we could not detect some of the 3-mm nodules.

No false-positive nodules were found compared with the control examinations. That is to say, all nodules detected on low-dose CT were actual nodular lesions. Nevertheless, in clinical practice, pulmonary parenchyma can be distorted and may contain scars and variations, which can erroneously be interpreted as a pulmonary nodule; thus the specificity in a clinical setting is usually decreased.

We found an increasing underestimation of the nodule volume at smaller nodule diameters, which was also found in some previous studies [15,16]. However, some other studies reported an increasing overestimation of the nodule volume at smaller nodule diameters [17-22]. For small pulmonary nodules, the transit zone between nodule and pulmonary background caused by partial volume effect was found to be important for accurate volumetry [23]. Thus, measurement errors in small nodules when measured manually should be considered.

In this study, we found that the measured nodule density was significantly different from the expected density. In lung cancer screening by unenhanced CT examination, accurate CT density is important mainly to differentiate between solid and nonsolid nodules, and to evaluate increase in density over time in the case of nonsolid nodules. However, as we only had one type of solid nodule, and two types of nonsolid nodule with a relatively

large difference in CT density compared with the solid nodules, no reliable conclusion can be drawn about the potential impact of CT density variation on lung cancer screening results. Future studies with more variation in the density of solid nodules, with CT densities within the clinically relevant range, have been planned to investigate the impact of CT density on nodule volumetry in more detail.

No difference was found between low-dose 16 and 64-row multi-detector CT from the same vendor regarding manual and semi-automated volumetry. However, a previous study found different nodule volumetry outcomes for four 16-row CT systems from different vendors [17]. As the follow-up of screened participants or clinical patients can last for an extensive period, different CT systems from different vendors may be used. A direct comparison of nodule volumes obtained from different CT systems from the same vendor seems valid, at least for the vendor investigated in this study. However, whether similar nodule volumes would have been found for other vendors is unknown.

Clinical implications

Some lung cancer screening projects were mainly based on nodule diameter [3], whereas other lung cancer screening projects were mainly based on nodule volume measurements [9]. In the National Lung Screening Trial (NLST) study, a positive result indicating suspected lung cancer on low-dose CT was defined as the presence of a nodule with a largest transverse axis of at least 4 mm [4]. In the Dutch-Belgian Randomized Lung Cancer Screening Trial (NELSON) study, a positive result was defined as either a fast-growing nodule with a volume of at least 50 mm^3, i.e., nearly 5 mm in diameter, or a nodule with a volume of at least 500 mm^3 [9]. The results of this study validate these screening protocols, as all solid nodules and nonsolid nodules with a diameter of 5 mm could be detected by observers on low-dose CT against an anthropomorphic pulmonary background.

Pulmonary nodule volumetry is used to guide the diagnostic strategy in the follow-up of lung cancer screening [9] . Repeated CT-derived volumetry of pulmonary nodules can be used to determine the risk of lung cancer, and can be used to monitor tumour response in the case of nonsurgical therapy. The accuracy and precision of pulmonary nodule volumetry depend on a number of factors, including image acquisition and reconstruction parameters, nodule characteristics, and the performance of algorithms for nodule segmentation and volume estimation [24]. Size measurement needs to be as accurate as possible in order to enable the assessment of nodule growth. A commonly used criterion for pulmonary nodule growth is given by the Response Evaluation Criteria in Solid Tumors (RECIST), which states that nodules in the stable disease category should not be larger than 20% or smaller than 30% in diameter on subsequent CT examinations [10]. However, a 20% error in diameter measurement could result in an error in volume for a spherical nodule of up to

73%, which could result in inaccurate growth rate evaluation. To improve accuracy in growth rate evaluation, semi-automated volumetry is favoured over manual volumetry.

Limitations

There are limitations to this study. Firstly, only spherical nodules were used with five discrete sizes. Additional data on the sensitivity of nodules with sizes between 3 and 5 mm in diameter is needed in order to determine the sensitivity of current low-dose CT and to optimise diagnostic screening strategies for small nodules. A further extension to our study is the assessment of the sensitivity of low-dose CT for nonspherical and irregular shaped (lobulated and/or spiculated) nodules. Secondly, we simulated healthy pulmonary tissues. The sensitivity of nodule detection is dependent on pulmonary structures and surrounding pathological lesions such as fibrosis, emphysema or consolidation could influence nodule detectability, which can make pulmonary nodules undetectable or be erroneously interpreted as pulmonary nodules. We therefore expect that the sensitivity in an in vivo setting may be lower, and that the false-positive rate may be higher, compared with the findings in this study. Thirdly, we used only one clinical low-dose CT screening protocol. Although the sensitivity of pulmonary nodules is also dependent on CT protocol, the protocol we used is the most common in current lung cancer screening projects using thin-slice, low-dose CT. Therefore, we expect that this protocol is the most relevant for the sensitivity of pulmonary nodule detection in lung cancer screening. Finally, semi-automated volumetry was not performed for nonsolid nodules, because the present commercially available volumetry software was only for solid nodules. In case of the considerable volumetry deviation from the actual volume in nonsolid nodules by manual measurements, a special software package for semi-automated volume measurement of nonsolid nodules is needed to assess these nonsolid nodules.

Conclusions

In conclusion, this anthropomorphic phantom study shows that a lung cancer screening protocol with low-dose CT is highly reliable for the detection of spherical pulmonary nodules of 5 mm in diameter and larger. Low-dose CT yields more accurate nodule volumetry when using a semi-automated software tool than manual measurements, with negligible underestimation of actual size, especially for small nodules. For early lung cancer detection, in which mostly small nodules are found, accurate measurement is especially necessary, to enable assessment of volume doubling time. Thus, a semi-automated volume measurement should be used in the setting of lung cancer screening. No difference in accuracy of volumetry was found between 16-row and 64-row multi-detector CT.

References

1. Shibuya K, Mathers CD, Boschi-Pinto C, Lopez AD, Murray CJ (2002) Global and regional estimates of cancer mortality and incidence by site: II. Results for the global burden of disease 2000. BMC Cancer 2:37.
2. Janssen-Heijnen ML, Coebergh JW (2001) Trends in incidence and prognosis of the histological subtypes of lung cancer in North America, Australia, New Zealand and Europe. Lung Cancer 31:123-137.
3. Aberle DR, Adams AM, Berg CD, et al. (2011) Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med 365:395-409.
4. Gatsonis CA, Natl Lung Screening Trial Res Team (2011) The National Lung Screening Trial: overview and study design. Radiology 258:243-253.
5. Nair A, Hansell DM (2011) European and North American lung cancer screening experience and implications for pulmonary nodule management. Eur Radiol 21:2445-2454.
6. Silverman JD, Paul NS, Siewerdsen JH (2009) Investigation of lung nodule detectability in low-dose 320-slice computed tomography. Med Phys 36:1700-1710.
7. Paul NS, Siewerdsen JH, Patsios D, Chung TB (2007) Investigating the low-dose limits of multidetector CT in lung nodule surveillance. Med Phys 34:3587-3595.
8. Ko JP, Rusinek H, Jacobs EL, et al. (2003) Small pulmonary nodules: volume measurement at chest CT - phantom study. Radiology 228:864-870.
9. van Klaveren RJ, Oudkerk M, Prokop M, et al. (2009) Management of lung nodules detected by volume CT scanning. The New England Journal of Medicine 361:2221-2229.
10. Eisenhauer EA, Therasse P, Bogaerts J, et al. (2009) New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer 45:228-247.
11. Higashihara H, Osuga K, Azuma T, et al. (2011) Detection of pulmonary nodules by C-arm CT using a phantom lung: comparison with CT. Acta Radiol 52:964-968.
12. Gavrielides MA, Kinnard LM, Myers KJ, et al. (2010) A resource for the assessment of lung nodule size estimation methods: database of thoracic CT scans of an anthropomorphic phantom. Opt Express 18:15244-15255.
13. Lee YC, Fullerton GD, Baiu C, Lescrenier MG, Goins BA (2011) Preclinical multimodality phantom design for quality assurance of tumor size measurement. BMC Med Phys 11:1.
14. Ravenel JG, Leue WM, Nietert PJ, Miller JV, Taylor KK, Silvestri GA (2008) Pulmonary nodule volume: effects of reconstruction parameters on automated measurements - a phantom study. Radiology 247:400-408.

15. Linning E, Ma D (2009) Volumetric measurement pulmonary ground-glass opacity nodules with multi-detector CT: effect of various tube current on measurement accuracy - a chest CT phantom study. Acad Radiol 16:934-939.
16. Zhang L, Yankelevitz DF, Henschke CI, Jirapatnakul AC, Reeves AP, Carter D (2010) Zone of transition: a potential source of error in tumor volume estimation. Radiology 256:633-639.
17. Das M, Ley-Zaporozhan J, Gietema HA, et al. (2007) Accuracy of automated volumetry of pulmonary nodules across different multislice CT scanners. Eur Radiol 17:1979-1984.
18. Kim KG, Goo JM, Kim JH, et al. (2005) Computer-aided diagnosis of localized ground-glass opacity in the lung at CT: Initial experience. Radiology 237:657-661.
19. Marten K, Funke M, Engelke C (2004) Flat panel detector-based volumetric CT - Prototype evaluation with volumetry of small artificial nodules in a pulmonary phantom. J Thorac Imaging 19:156-163.
20. Tao P, Griess F, Lvov Y, et al. (2004) Characterization of small nodules by automatic segmentation of X-ray computed tomography images. J Comput Assist Tomogr 28:372-377.
21. Way TW, Chan HP, Goodsitt MM, et al. (2008) Effect of CT scanning parameters on volumetric measurements of pulmonary nodules by 3D active contour segmentation: a phantom study. Phys Med Biol 53:1295-1312.
22. Winer-Muram HT, Jennings SG, Meyer CA, et al. (2003) Effect of varying CT section width on volumetric measurement of lung tumors and application of compensatory equations. Radiology 229:184-194.
23. Goo JM, Tongdee T, Tongdee R, Yeo K, Hildebolt CF, Bae KT (2005) Volumetric measurement of synthetic lung nodules with multi-detector row CT: effect of various image reconstruction parameters and segmentation thresholds on measurement accuracy. Radiology 235:850-856.
24. Gavrielides MA, Kinnard LM, Myers KJ, Petrick N (2009) Noncalcified lung nodules: volumetric assessment with thoracic CT. Radiology 251:26-37.

Chapter 4

Small Irregular Pulmonary Nodules in 64-Row Multi-Detector CT: Observer Detection Sensitivity and Volumetry Accuracy

Accepted by American Journal of Roentgenology

Xueqian Xie[12] / Martin J. Willemink[3] / Pim A. de Jong[3] / Peter M.A. van Ooijen[12]
Matthijs Oudkerk[2] / Rozemarijn Vliegenthart[12] / Marcel J.W. Greuter[1]

University of Groningen, University Medical Center Groningen,
[1]Department of Radiology, [2]Center for Medical Imaging – North East Netherlands;
[3]Department of Radiology, University Medical Center Utrecht

Abstract

Objective

To evaluate observer detection and volume measurement of small irregular solid artificial pulmonary nodules on 64-row multi-detector computed tomography (CT) in an anthropomorphic thoracic phantom.

Materials and Methods

Forty in-house made solid pulmonary nodules (lobulated and spiculated; actual volume 5.1 to 88.4 mm^3; actual CT densities -51 to +157 HU) were randomly placed inside an anthropomorphic thoracic phantom with pulmonary vasculature. The phantom was examined on two 64-row multi-detector CT scanners, using a scan protocol as applied in lung cancer screening. Two independent blinded observers screened for pulmonary nodules. Nodule volume was evaluated semi-automatically using dedicated software, and compared to the actual volume using a t-test. The inter-scanner and inter-observer agreement of volumetry was assessed using Bland-Altman analysis.

Results

Observer detection sensitivity increased with raising size of irregular nodules. Sensitivity was 100% when actual volume was at least 69 mm^3, regardless of specific observer, scanner, nodule shape and density. Overall, nodule volume was underestimated by 18.9 ± 11.8 mm^3 (39 ± 21%, $p < 0.001$). Relative inter-scanner difference of volumetry was 3.3% (95% CI: -33.9% to 40.4%). Relative inter-observer difference was 0.6% (-33.3% to 34.5%).

Conclusions

Small irregular solid pulmonary nodules with an actual volume of at least 69 mm^3 are reliably detected on 64-row multi-detector CT. However, CT-derived volume of those small nodules is largely underestimated, with considerable variation.

Introduction

Lung cancer is worldwide the most common cause for cancer death, and accounts for more than 18% of the total deaths from cancer [1]. Since lung cancer is predominantly found at a relatively late stage, its 5-year survival is only 15% or even less [2]. The National Lung Screening Trial (NLST) has demonstrated a promising result that early detection of lung cancer with computed tomography (CT) reduces mortality [3]. Therefore, CT allows the potential to be an effective screening tool for early detection and mortality reduction of lung cancer [4,5].

In the routine clinical population, there is inconsistency whether surveillance CT should be performed after encountering an overwhelming number of indeterminate small solid pulmonary nodules on CT examinations, as small as < 6 mm in diameter (approximately 90 mm^3) [6]. In a high risk population, The Fleischner Society recommends follow-up CT for those small nodules, because small nodules have a nonignorable opportunity to develop into malignancy [7]. The common participants in lung cancer screening trials are elderly and smokers, two major risk factors for lung cancer [3,5]. Optimal management protocol of small pulmonary nodules is important for early distinction of malignant lesions in screening.

Assessment of small nodules starts with sensitive observer detection and accurate growth evaluation. Computed assisted diagnosis (CAD) or observer consensus has been used as a reference for pulmonary nodule detectability in human studies [8-10]. To reach more precise nodule evaluation, number and volume of the actually existing nodules need to be certain. Thus, an anthropomorphic phantom study is necessary, based on artificial nodules with a known number and volume. In-depth data from well-designed phantom studies is limited for nodules less than 6 mm in diameter, which are essential for optimization of nodule management protocol for those small nodules in lung cancer screening [11-13]. As an extension of our previous phantom study regarding focusing on spherical nodules [11], the current study aims at irregular nodules. The purpose is to evaluate the detection and volumetry of small irregular solid pulmonary nodules, which were randomly placed in an anthropomorphic phantom with pulmonary vasculature.

Materials and Methods

Phantom

An anthropomorphic chest phantom (Lungman, Kyoto Kagaku, Tokyo, Japan) was used, with an artificial thoracic wall, nonbeating heart, mediastinum, spine, ribs, diaphragm and lungs with pulmonary vessels (Figure 1). The phantom was an accurate life-size anatomical model of a healthy male thorax. Soft tissues were imitated with materials made of polyure-

thane resin composites and bones were imitated with materials made of epoxy resin. X-ray absorption rates of these materials were very close to those of human tissues. The thoracic cavity consisted of pulmonary vessels surrounded by air.

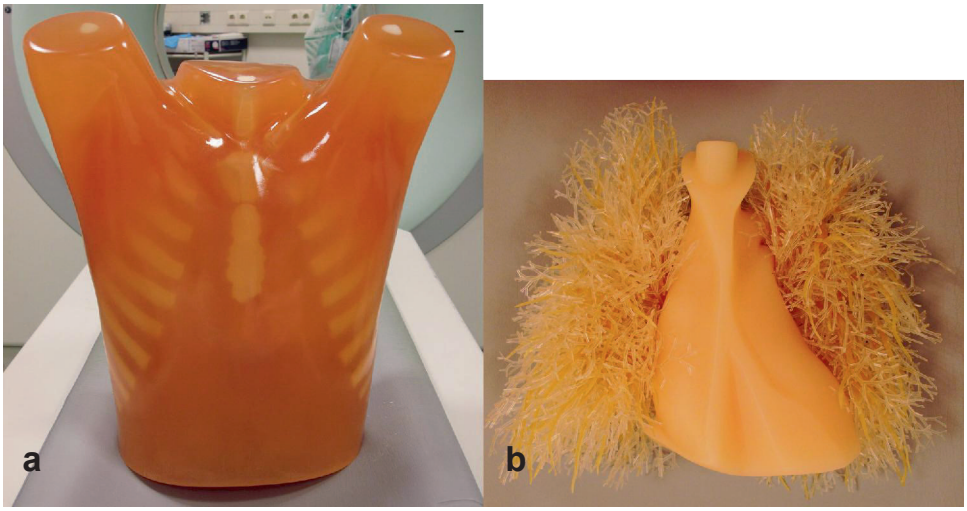


Figure 1 Photograph of the anthropomorphic chest phantom (**a**) and of the internal artificial heart and lungs (**b**).

Preparation of artificial pulmonary nodules

Forty artificial pulmonary nodules were made in-house, with two irregular shapes (spiculated and lobulated) and five actual CT densities (-51, +2, +57, +125 and +157 Hounsfield Units (HU) at 120 kV). A three-dimensional printer (Objet Eden 250, Objet, Rheinmünster, Germany) was used to make the nodules of +157 HU, and the molds to prepare the nodules of the other four CT densities. The artificial nodules of +157 HU were made of plastic (Verowhite 830, Objet, Rheinmünster, Germany) with known physical density. The nodules of the other four CT densities were made of paraffin mixed with contrast media (Lipiodol 480 mgI/ml, Guerbet, Villepinte, France).

To make the artificial nodules, firstly, we prepared and scanned large pieces of paraffin mixed with contrast media. The proportion between those two components was adjusted, until the pre-defined actual CT density was reached. Volume (v_1) of these large pieces of paraffin was measured by submersion into a measuring cylinder filled with water. Mass (m_1) was measured by an accurate balance. Density (ρ) was calculated by $\rho = m_1 / v_1$. Secondly, paraffin mixed with contrast media was melted and poured into molds to shape the nodules. Finally, 40 artificial small irregular nodules were made. Mass (m_2) of each nodule was weighted using the balance. Actual volume (v_2) of each artificial nodule was

calculated by $v_2 = m_2 / \rho$. These measurements were repeated three times. Photographs and actual volumes of the 40 artificial nodules were shown in Table 1. The actual volume of the simulated nodules ranged from 5.1 ± 0.2 mm^3 to 88.4 ± 3.4 mm^3, with a mean volume of 33 mm^3.

CT imaging

Two 64-row multi-detector CT scanners were utilized (CT-A: Sensation 64, Siemens Healthcare, Forchheim, Germany; CT-B: Brilliance 64, Philips Healthcare, Best, The Netherlands), using a clinically applied CT acquisition protocol for lung cancer screening [13]. The image acquisition protocol for CT-A was: spiral acquisition at 120 kV, 20 mAs, rotation time 0.5 s, pitch 1.5, collimation 2 × 32 × 0.6 mm, field of view 300 mm, volume CT dose index ($CTDI_{vol}$) 1.6 mGy and dose length product (DLP) 48 mGy·cm, without using tube current modulation. The CT images were reconstructed at a 1.0 / 0.7 mm slice thickness / increment using a medium-smooth B30f kernel. The image acquisition protocol for CT-B was: spiral acquisition at 120 kV, 20 mAs, rotation time 0.5 s, pitch 1.39, collimation 64 × 0.625 mm, field of view 300 mm, $CTDI_{vol}$ 1.3 mGy and DLP 39 mGy·cm, without utilizing tube current modulation. The CT images were reconstructed at a 1.0 / 0.7 mm slice thickness / increment using a medium-smooth B kernel.

The thoracic phantom was examined once without any pulmonary nodules to serve as a control examination, in order to determine possible false positive nodules. Furthermore, all nodules were examined on the CT table without the phantom, so as to confirm the visibility on CT.

Artificial nodules were randomly positioned in the anthropomorphic lungs. All nodules were attached to pulmonary vessels. None of those nodules was connected to pleura or located sub-pleurally. A randomly defined set from 0 to 9 nodules was positioned inside the anthropomorphic phantom in an examination. Nodule location was randomly determined. Each nodule out of 40 artificial nodules was placed five times in different locations. Thus, per CT scanner 51 examinations with nodules were scheduled. Each examination was subsequently repeated three times. The position of human subjects is commonly not perfectly parallel to the central axis of the CT examination bed (z-axis). We simulated this by rotation of the phantom 5 - 10 degrees between each repeated examination. Also, the location of individuals on the CT examination bed commonly changes between examinations. We simulated this by translation of the phantom 5 - 10 cm between each repeated examination. In total, per CT scanner 153 examinations were performed, in which 600 nodule images were acquired.

Table 1 Volume measurements of the 40 small artificial irregular pulmonary nodules

Actual density ,HU	Spiculated nodules			Lobulated nodules		
	Photograph	Actual volume, mm^3	CT measured volume, mm^3	Photograph	Actual volume, mm^3	CT measured volume, mm^3
-51±11		5.9±0.3	n/d		10.6±0.5	6.7±1.3*
		25.6±1.2	10.5±3.2*		23.1±1.1	11.6±1.9*
		31.8±1.5	12.4±7.2*		34.9±1.7	17.7±4.8*
		45.9±2.2	17.7±5.0*		74.5±3.6	47.0±9.5*
+2±13		15.4±0.6	n/d		13.6±0.6	9.0±1.2*
		19.1±0.8	14.2†		25.5±1.0	13.8±6.7*
		37.5±1.4	25.1±4.7*		51.1±2.0	26.7±6.7*
		43.9±1.7	20±11*		88.4±3.4	51.4±3.9*
+57±13		18.3±0.2	12.6±6.1*		13.0±0.1	7.7±2.8*
		34.8±0.3	14.0±4.8*		17.0±0.2	7.9±0.6*
		35.3±0.7	18.1±6.0*		41.9±0.1	26.2±6.9*
		47.1±1.0	29.1±5.8*		80.1±0.1	50.6±6.7*
+125±20		20.6±0.7	3.8±1.1*		10.2±0.4	6.4†
		23.8±0.9	11.5±2.5*		15.4±0.6	11.1±2.5*
		49.5±1.8	20±10*		33.0±1.2	20.1±7.8*
		68.9±2.5	38.6±8.8*		54.9±2.0	45.5±7.8*
+157±8		5.1±0.2	n/d		7.6±0.3	n/d
		12.0±0.4	5.8±1.7*		15.9±0.6	13.3±6.1*
		23.4±0.8	15.5±6.4*		33.8±1.2	27.8±6.8*
		40.9±1.4	25.9±5.9*		72.7±2.5	59.1±5.1*

*: $p < 0.01$ when compared with the actual volume.

†: Because of insufficient sample size, standard deviation and statistical significance were not calculated. n/d = not detectable.

Image analysis

Two independent observers evaluated the examinations, including a radiologist (observer 1) with 9 years of experiences in diagnostic thoracic imaging, and a resident (observer 2) with two years of radiological experiences. Each observer independently read all examinations of two CT scanners, using a commercial workstation (Somaris/5 syngo, Siemens, Forchheim, Germany). Both observers were blinded to information about the presence, number and location of the artificial nodules.

The observers reviewed the images on maximum intensity projection (MIP) and thin-slice axial and coronal display, using a dedicated software tool (LungCARE, Siemens, Forchheim, Germany). The slab thickness of MIP display was 6 mm. The observers were instructed to review the images for the presence of nodules within clinically representative time duration of approximately two minutes per examination. When a nodule image was found in an examination, the observer documented the number and location of the observed nodule image. Next, the nodule image was compared to the control examination (without a nodule), to verify whether the finding was true positive (TP) or false positive. If an observed TP nodule was segmentable, CT-derived volume was semi-automatically measured.

Statistical analysis

Results are presented as mean ± standard deviation (SD). Observer detection sensitivity was calculated per nodule, as a percentage of the TP findings. The relationship between sensitivity and actual nodule volume was explored using curve fitting based on a sigmoid function model [14]. The influential effects on nodule evaluation were assessed by univariate analysis with a general linear model. In that univariate analysis, dependent variable was observer sensitivity or CT-derived volume. Independent factors were observer (observer 1 and 2), CT scanner (CT-A and CT-B), nodule shape (spiculated and lobulated), actual volume (5.1 to 88.4 mm^3) and density (-51, +2, +57, +125 and +157 HU).

The inter-scanner and inter-observer reliability for CT-derived volume was expressed using intra-class correlation coefficients (ICC). ICC values larger than 0.90 were ranked as high agreement, in the range of 0.75 to 0.90 were rated as moderate, and smaller than 0.75 were considered as low [15, 16]. The inter-scanner and inter-observer agreement of nodule volumetry was evaluated using Bland-Altman analysis. A 95% confidence interval (CI) was expressed as mean ± 1.96 SD.

The difference in CT-derived and actual volume was evaluated by an independent-samples t-test. The percentage deviation of CT-derived volume from the actual volume was calculated as (CT-derived volume - actual volume) / actual volume × 100%.

A p-value below 0.05 is considered statistically significant. Statistical analysis was performed using SPSS version 18.0 (IBM, New York, USA) and SigmaPlot version 12.1 (Systat Software, San Jose, USA).

Results

Observer detection sensitivity

When scanned outside the phantom on the CT table, all the 40 nodules could be visualized on CT regardless of nodule size. However, within the phantom, the first observer detected 340 (56.7%) and 315 (52.5%) TP nodule images out of the maximum 600 nodule images in CT-A and CT-B, respectively. The second observer found 245 (40.8%) and 261 (43.5%), respectively. In addition, the first observer identified 3 (0.5%) false positive nodule images in CT-A. The second observer identified 2 (0.3%) in CT-A and 2 (0.3%) in CT-B. Representative CT images of the anthropomorphic thoracic phantom and nodules are shown in Figure 2.

Table 2 Influential effects on nodule evaluation, assessed by univariate analysis

Influential factor	Observer sensitivity		CT-derived volume	
	Beta	p value	Beta	p value
Observer	-0.17	<0.05	0.06	0.17
Scanner	-0.01	0.90	-0.08	<0.01
Shape	0.11	0.16	0.43	<0.001
Actual volume	0.10	<0.001	0.84	<0.001
Actual density	0.16	0.31	0.01	0.73

Dependent variable was observer sensitivity or CT-derived volume. Independent factors were observer (observer 1 and 2), CT scanner (CT-A and CT-B), nodule shape (spiculated and lobulated), actual volume (5.1 to 88.4 mm^3) and density (-51, +2, +57, +125 and +157 HU).

Observer sensitivity was positively associated with the first observer and larger actual volume ($p < 0.05$). But scanner, nodule shape and density were not significant factors ($p > 0.05$) (Table 2). Therefore, sensitivity data were combined for scanner, nodule shape, and density.

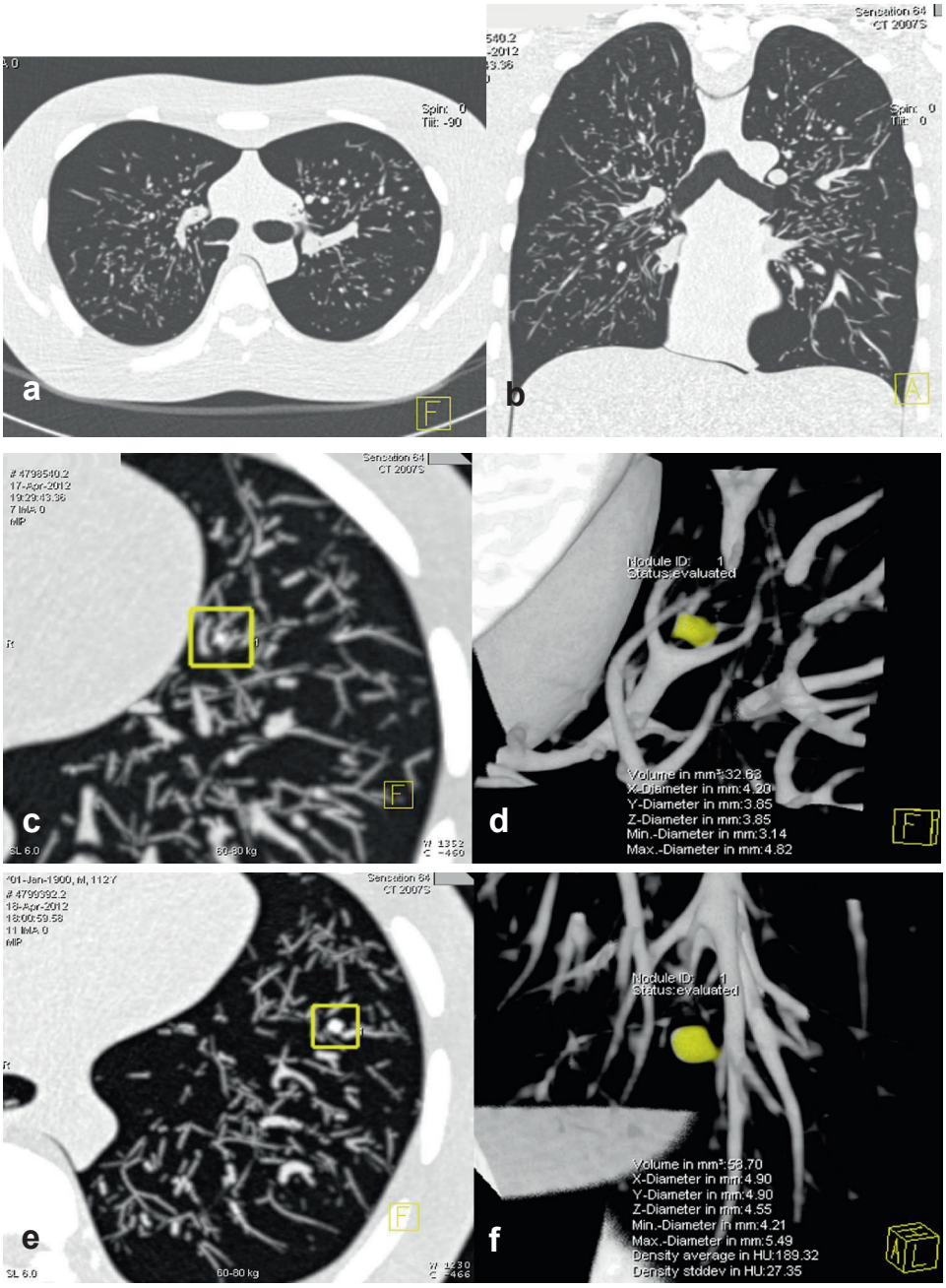
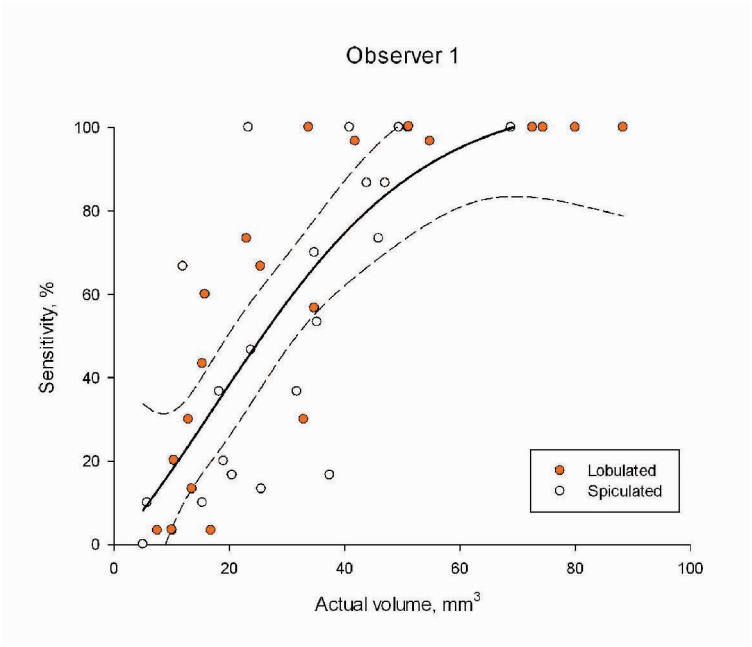


Figure 2 Axial (**a**) and coronal (**b**) computed tomographic image of the anthropomorphic chest phantom. Maximal intensity projection (**c**, MIP) and three-dimensional (**d**, 3D) reconstruction image of an artificial spiculated nodule. MIP (**e**) and 3D image (**f**) of an artificial lobulated nodule.



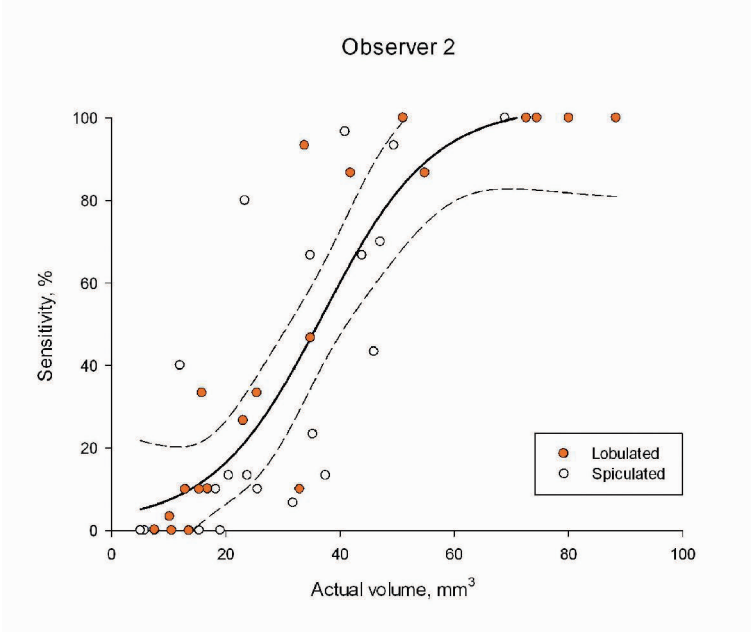


Figure 3 Observer detection sensitivity of artificial small irregular pulmonary nodules by observer 1 (**a**) and observer 2 (**b**). The sensitivity fitted a sigmoid curve ($p < 0.001$) for both observers (solid lines), with 95% confidence bands indicated (dashed lines). The R values of the curve fitting were 0.80 and 0.86 for observer 1 and 2, respectively.

Observer detection sensitivity fitted a sigmoid curve, versus actual nodule volume ($p < 0.001$) for both observers (Figure 3). Those curves demonstrated that sensitivity increased with larger nodule volume. In case of nodules with an actual volume < 15 mm^3, sensitivity was 6% to 28% of the first observer, and 5% to 11% of the second. For nodules <50 mm^3, sensitivity of the first observer was slightly higher than of the second. When actual volume increased to ≥ 69 ± 2 mm^3 (CT-derived volume 39 ± 6 mm^3, 95% CI: 27 to 51 mm^3), the sensitivity was always 100%, regardless of observer, scanner, nodule shape and density.

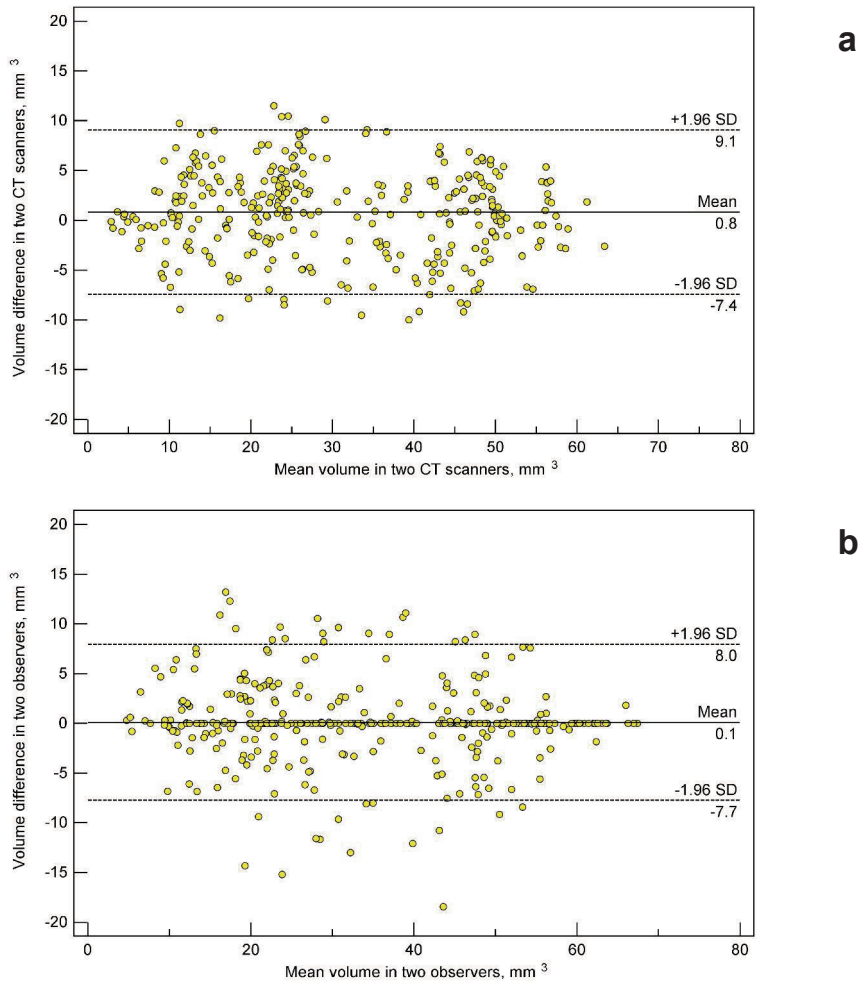


Figure 4 Bland-Altman plots for inter-scanner (**a**) and inter-observer agreement (**b**) of CT-derived volume. The agreement is expressed as relative difference, with 95% confidence bands indicated as dashed lines.

Nodule volumetry

Majority of the observed nodules were successfully segmented and measured. The first observer evaluated 297 (87.4%) out of 340 TP nodule images and 226 (71.7%) out of 315, in CT-A and CT-B, respectively. The second observer evaluated 227 (92.7%) out of 245 and 207 (79.3%) out of 261, in CT-A and CT-B, respectively. The nonsegmentable nodules were not measured. Actual nodule volume, CT scanner and nodule shape significantly influenced CT-derived volume ($p < 0.01$). Yet observer and actual density were not significant factors ($p > 0.05$) (Table 2).

Inter-scanner and inter-observer reliability of CT-derived volume was high, with an ICC of 0.90 ($p < 0.001$) and 0.95 ($p < 0.001$), respectively. In Bland-Altman analysis, the relative difference of volumetry between two CT scanners was 3.3% (95% CI: -33.9% to 40.4%). The relative difference between two observers was 0.6 % (-33.3% to 34.5%) (Figure 4).

CT-derived volume of the segmentable nodules was significantly lower than actual volume ($p < 0.01$) (Table 1). The mean underestimation was 18.9 ± 11.8 mm^3 (percentage underestimation: 39 ± 21%). CT-derived volume in CT-A and CT-B was underestimated by 18.5 ± 12.0 mm^3 (38 ± 21%) and 19.3 ± 10.8 mm^3 (40 ± 15%), respectively. CT-derived volume of the spiculated and lobulated nodules was underestimated by 21.0 ± 11.2 mm^3 (48 ± 21%) and 17.2 ± 12.2 mm^3 (33 ± 19%), respectively.

Discussion

In this anthropomorphic phantom study for small irregular solid pulmonary nodules, observer detection sensitivity increased with raising nodule volume. When actual volume was at least 69 mm^3 (CT-derived volume 39 ± 6 mm^3), sensitivity was always 100%. The 95% CI of relative inter-scanner and inter-observer volumetry difference is within 40%. CT-derived volume was considerably underestimated by 39%, compared to the actual volume.

In some previous phantom studies on nodule detectability, spherical nodules with an actual volume of 4 to 15 mm^3 were all detected [17,18]. Yet the nodules were placed in a known order and examined inside a phantom without pulmonary vasculature. Using an anthropomorphic phantom with pulmonary vessels, detection sensitivity fell to 60% to 80% for randomly positioned spherical nodules of 14 mm^3 [11]. In our current study, as an extension to the aforementioned study using the same anthropomorphic phantom, sensitivity diminished to ≤ 28% for randomly positioned irregular nodules of less than 15 mm^3. In this well-controlled in-vitro setting, other influential factors were excluded. Pulmonary vasculature and irregular nodule shape contribute to explain the decreased observer sensitivity in case of small irregular nodules.

We found that larger nodule volume resulted in higher observer sensitivity. When nodules were larger than 69 mm^3 in actual volume, all the nodules were detected. This finding was similar to the previous study on spherical nodules, in which nodules larger than 65 mm^3 were all detected [11]. Thus, the current study strengthened the finding of nodule detection sensitivity. Pulmonary nodules larger than 69 mm^3 can be detected with high reliability, regardless of specific observer, CT scanner and nodule characteristics. However, nodules smaller than 65 mm^3 cannot be reliably detected by observers.

Only a few false-positive nodules were found (0% to 0.5%), due to erroneous interpretation of overlapping vessels as nodules. In-vivo pulmonary structures contain scars and variations, which could lead to false-positive interpretations. Thus, one would expect higher false-positive rates in a clinical setting, compared to 0% to 0.5% in the present study.

Variability has been found in semi-automatically CT-derived nodule volumetry [19, 20]. An in-vivo study showed that the 95% CI of inter-scan variability of nodule volumetry for solid nodules from 15 to 500 mm^3 was within 25%, based on two CT examinations performed on the same day [21]. In our study, the 95% CI of relative inter-scanner and inter-observer difference was within 40%, for irregular nodules smaller than 90 mm^3. Variability of volumetry seems larger in small irregular nodules.

CT-derived volume of small irregular solid nodules was largely underestimated by 39% in this study. In the previous study on spherical nodules of similar size, CT-derived volume was underestimated by ≤ 25% [11,22]. In contrast, CT-derived volume was overestimated in some publications based on larger spherical and irregular nodules [23-25]. It is well known that volumetry accuracy of pulmonary nodules depends on a number of factors, e.g., image processing and nodule characteristics [26,27]. In CT, there is transition in Hounsfield values between high and low density objects, caused by the partial volume effects [28]. In our study, this transition is between a nodule and the surrounding pulmomary parenchyma, which degrades the exactness of nodule segmentation, especially for small nodules [29, 30]. The transition around an irregular nodule was expected to be larger than around a spherical nodule with a smooth contour. Consequently, small irregular nodules showed a larger underestimation of CT-derived volume than spherical nodules.

Several nodule characteristics were fabricated in this study. Firstly, irregular nodules were frequent findings in lung cancer screening [19,31]. Secondly, solid nodules were the majority of observed nodules, rather than nonsolid nodules [5]. Thirdly, CT density range from -57 to 157 HU was usual in solid nodules [31]. Finally, nodule number range from 0 to 9 per participant was also common [32]. Therefore, we expected that these characteristics were closely mimics to the real nodules in lung cancer screening.

Clinical implications

Different lower limits of CT-derived nodule volume were utilized in lung cancer screening trials to define a nodule in which malignancy cannot be excluded, leading to follow-up examination or further work-up, for example, 50 mm^3 in the Dutch-Belgian Randomized Lung Cancer Screening Trial, 60 mm^3 in the Multi-centric Italian Lung Detection trial and 4 mm in diameter (approximately 35 mm^3 in volume) in the NLST [5, 33]. Of course, before nodule management can be determined, the nodule first has to be detected. Our results show that nodules with CT-derived volume of 39 ± 6 mm^3 (95%CI: 27 to 51 mm^3) can be reliably detected by observers, at least in a phantom setting, independent of CT scanner and nodule characteristics. This finding supports the use of a nodule CT volume cut-off of about 50 mm^3 to determine presence of a nodule, as nodules of this CT size are not likely to be missed.

The 95% CI of variability in nodule volumetry is used to exclude systematic errors, and to define a stable nodule. A volume increase of at least 25% has been defined as the minimum to distinguish growth from measurement variability, for small and indeterminate nodules with CT volume < 500 mm^3 [5,21,33]. Because it is unsure whether a CT-derived nodule growth < 25% is caused by true growth or by measurement variability, lung cancer screening subjects with CT-derived nodule growth < 25% will be examined in another follow-up examination to confirm nodule stability or reliably detect nodule growth, instead of immediate action [4,5]. In our study, the 95% CI of variability was within 40% for nodules < 90 mm^3 (95% CI of CT derived volume, 44 to 60 mm^3). This means that, because of higher variability, we cannot distinguish true growth from measurement variability when a nodule <90 mm^3 increases in volume less than 40%. Here, using the 25% cut-off criterion would lead to a higher number of false-positive results, potentially resulting in avoidable patient anxiety and morbidity. Another follow-up examination seems appropriate when a nodule < 90 mm^3 increases in volume less than 40%, to confirm stability or reliably detect growth.

If the measured volume of a nodule is underestimated in two repeated examinations but to a similar extent (percentage wise), the calculated nodule growth rate is still accurate. Thus, nodule management of intermediate size nodules based on repeated CT examinations and assessment of volume doubling time, such as in the NELSON study [34], is not affected.

Small irregular nodules yield a larger underestimation in CT-derived volumetry than small spherical nodules (underestimation by 39% versus 20% compared to actual volume) [11]. This may affect the assessment of nodule management. E.g., a nodule with actual size of 60 mm^3 could be measured in CT as 50 mm^3 in case of spherical shape, but 40 mm^3 in case of irregular shape. Thus, in the latter case, the nodule would fall in the small size category (< 50 mm^3), which receives standard follow-up examination, instead of short-term

repeat CT scanning [34]. Although this could have resulted in misclassification of indeterminate size, irregular nodules at the baseline round of the NELSON study, in view of the very low interval cancers [5], it is unlikely that this potential misclassification has had clinical impact. Change in nodule shape between CT examinations could affect volumetry and thus, assessment of growth rate. The implication of this finding could be explained by assuming the following scenario. If there is an indeterminate spherical nodule of an actual volume of 63 mm^3, the volume is generally underestimated in CT by 20%, resulting in a CT-derived volume of 50 mm^3. When that nodule is followed up and examined again after 90 days, it might have developed into an irregular nodule of 81 mm^3 in actual volume. This corresponds to a volume doubling time of 315 days, and thus, the nodule should be assessed as fast growing [34], and suspected for malignancy. As, in case of irregular nodules, underestimation of actual volume is generally larger and can go up to 39% it is possible that the CT-derived volume remains around 50 mm^3. Thus, in this particular scenario, misdiagnosis of that nodule would have occurred. This scenario indicates that when nodule shape is not considered in case of small nodules, there is a chance to misjudge growth rate. Therefore, since a pulmonary nodule has an opportunity to develop into another appearance during growth [6, 35], volumetry errors as a function to nodule shape have to be considered during follow-up, at least as demonstrated in our phantom studies for nodules smaller than 90 mm^3.

Limitations

Firstly, only 40 solid nodules were simulated in this study, which is inconsistent with generally varying shapes in real patients. We expected that those artificial nodules were representative for small irregular pulmonary nodules. Secondly, healthy pulmonary tissues were simulated in this study. Nodule detection depends on successful distinction of a pulmonary nodule from normal pulmonary structures and surrounding pathological lesions, such as fibrosis and consolidation, which make pulmonary nodules undetectable or lead to erroneous interpretation. Hence, sensitivity might be lower in real patients, and false positive rates might be higher. Thirdly, we only tested one CT acquisition protocol and dedicated volumetry software. That protocol has been widely applied in lung cancer screening [4]. That software has also been frequently utilized in screening [5,36,37]. However, accuracy of nodule volumetry depends on acquisition and algorithm in individual settings [27,38]. When discussing the results, we linked to the studies based on similar CT acquisition protocol and the same software [11,19,21,31,35]. Thus, our results might be comparable to those previous data. A further extension is the evaluation in different acquisition and algorithm environments.

Conclusions

This study adds understandings for evaluation of irregular solid pulmonary nodules, smaller than 6 mm in diameter (approximately 90 mm^3) in lung cancer CT screening. Observers reliably detect irregular nodules with an actual volume of at least 69 mm^3 (CT-derived volume 39 ± 6 mm^3), regardless of specific CT scanner, observer and nodule characteristics. Relative inter-scanner and inter-observer difference of volumetry is within 40%. Irregular nodules yield a larger underestimation in CT-derived volumetry than spherical nodules.

References

1. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer 2010; 127:2893-2917
2. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. CA: a cancer journal for clinicians 2012; 62:10-29
3. Aberle DR, Adams AM, Berg CD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med 2011; 365:395-409
4. Nair A, Hansell DM. European and North American lung cancer screening experience and implications for pulmonary nodule management. Eur Radiol 2011; 21:2445-2454
5. van Klaveren RJ, Oudkerk M, Prokop M, et al. Management of lung nodules detected by volume CT scanning. N Engl J Med 2009; 361:2221-2229
6. Beigelman-Aubry C, Hill C, Grenier PA. Management of an incidentally discovered pulmonary nodule. Eur Radiol 2007; 17:449-466
7. MacMahon H, Austin JHM, Gamsu G, et al. Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner Society. Radiology 2005; 237:395-400
8. Yingru Z, de Bock GH, Vliegenthart R, et al. Performance of computer-aided detection of pulmonary nodules in low-dose CT: comparison with double reading by nodule volume. European Radiology 2012; 22:2076-2084
9. Armato SG, III, Roberts RY, Kocherginsky M, et al. Assessment of radiologist performance in the detection of lung nodules: dependence on the definition of "truth". Academic Radiology 2009; 16:28-38
10. Marten K, Engelke C. Computer-aided detection and automated CT volumetry of pulmonary nodules. European Radiology 2007; 17:888-901
11. Xie X, Zhao Y, Snijder R, et al. Sensitivity and accuracy of volumetry of pulmonary nodules on low-dose 16- and 64-row multi-detector CT: an anthropomorphic phantom study. Eur Radiol 2013; 23:139-147
12. Nietert PJ, Ravenel JG, Leue WM, et al. Imprecision in Automated Volume Measurements of Pulmonary Nodules and Its Effect on the Level of Uncertainty in Volume Doubling Time Estimation. Chest 2009; 135:1580-1587
13. Ru Zhao Y, Xie X, de Koning HJ, Mali WP, Vliegenthart R, Oudkerk M. NELSON lung cancer screening study. Cancer imaging 2011; 11 S79-84
14. von Seggern D. CRC standard curves and surfaces with mathematics, 2nd ed. Boca Raton, FL: CRC Press 2007;
15. Lundberg M, Grimby-Ekman A, Verbunt J, Simmonds MJ. Pain-related fear: a critical review of the related measures. Pain research and treatment 2011; 2011:494196

16. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33:159-174
17. Ko JP, Rusinek H, Jacobs EL, et al. Small pulmonary nodules: volume measurement at chest CT - phantom study. Radiology 2003; 228:864-870
18. Lee YC, Fullerton GD, Baiu C, Lescrenier MG, Goins BA. Preclinical multimodality phantom design for quality assurance of tumor size measurement. BMC Med Phys 2011; 11:1
19. Wang Y, van Klaveren RJ, van der Zaag-Loonen HJ, et al. Effect of nodule characteristics on variability of semiautomated volume measurements in pulmonary nodules detected in a lung cancer screening program. Radiology 2008; 248:625-631
20. Willemink M, Leiner T, Budde R, et al. Systematic Error in Lung Nodule Volumetry: Effect of Iterative Reconstruction Versus Filtered Back Projection at Different CT Parameters. Am J Roentgenol 2012; 199:1241-1246
21. Gietema HA, Schaefer-Prokop CM, Mali WPTM, Groenewegen G, Prokop M. Pulmonary nodules: Interscan variability of semiautomated volume measurements with multisection CT-influence of inspiration level, nodule size, and segmentation performance. Radiology 2007; 245:888-894
22. Prionas ND, Ray S, Boone JM. Volume assessment accuracy in computed tomography: a phantom study. J Appl Clin Med Phys 2010; 11:3037
23. Baiyu C, Barnhart H, Richard S, Colsher J, Amurao M, Samei E. Quantitative CT: technique dependence of volume estimation on pulmonary nodules. Physics in Medicine and Biology 2012; 57:1335-13481348
24. Das M, Ley-Zaporozhan J, Gietema HA, et al. Accuracy of automated volumetry of pulmonary nodules across different multislice CT scanners. European Radiology 2007; 17:1979-1984
25. Marten K, Dullin C, Machann W, et al. Comparison of flat-panel-detector-based CT and multidetector-row CT in automated volumetry of pulmonary nodules using an anthropomorphic chest phantom. British Journal of Radiology 2009; 82:716-723
26. de Jong PA, Leiner T, Lammers JWJ, Gietema HA. Can Low-Dose Unenhanced Chest CT Be Used for Follow-Up of Lung Nodules? American Journal of Roentgenology 2012; 199:777-780
27. Christe A, Torrente JC, Lin M, et al. CT Screening and Follow-Up of Lung Nodules: Effects of Tube Current-Time Setting and Nodule Size and Density on Detectability and of Tube Current-Time Setting on Apparent Size. American Journal of Roentgenology 2011; 197:623-630
28. Groen JM, Greuter MJ, Vliegenthart R, et al. Calcium scoring using 64-slice MDCT, dual source CT and EBT: a comparative phantom study. Int J Cardiovasc Imaging 2008; 24:547-556

29. Goo JM, Tongdee T, Tongdee R, Yeo K, Hildebolt CF, Bae KT. Volumetric measurement of synthetic lung nodules with multi-detector row CT: effect of various image reconstruction parameters and segmentation thresholds on measurement accuracy. Radiology 2005; 235:850-856
30. Wiemker R, Rogalla P, Blaffert T, et al. Aspects of computer-aided detection (CAD) and volumetry of pulmonary nodules using multislice CT. British Journal of Radiology 2005; 78:S46-S56
31. Xu DM, van Klaveren RJ, de Bock GH, et al. Limited value of shape, margin and CT density in the discrimination between benign and malignant screen detected solid pulmonary nodules of the NELSON trial. European Journal of Radiology 2008; 68:347-352
32. Wang Y, de Bock GH, van Klaveren RJ, et al. Volumetric measurement of pulmonary nodules at low-dose chest CT: effect of reconstruction setting on measurement variability. European Radiology 2010; 20:1180-1187
33. Marchiano A, Calabro E, Civelli E, et al. Pulmonary nodules: volume repeatability at multidetector CT lung cancer screening. Radiology 2009; 251:919-925
34. Xu DM, Gietema H, de Koning H, et al. Nodule management protocol of the NELSON randomised lung cancer screening trial. Lung Cancer 2006; 54:177-184
35. Yankelevitz DF, Gupta R, Zhao B, Henschke CI. Small pulmonary nodules: evaluation with repeat CT--preliminary experience. Radiology 1999; 212:561-566
36. Pastorino U, Rossi M, Rosato V, et al. Annual or biennial CT screening versus observation in heavy smokers: 5-year results of the MILD trial. Eur J Cancer Prev 2012; 21:308-315
37. Lopes Pegna A, Picozzi G, Mascalchi M, et al. Design, recruitment and baseline results of the ITALUNG trial for lung cancer screening with low-dose CT. Lung Cancer 2009; 64:34-40
38. Willemink MJ, Borstlap J, Takx RAP, et al. The effects of computed tomography with iterative reconstruction on solid pulmonary nodule volume quantification. PloS one 2013; 8:e58053

Chapter 5

Inter- and Intra-scanner Variability of Pulmonary Nodule Volumetry on Low-Dose 64-Row CT: an Anthropomorphic Phantom Study

British Journal of Radiology. 2013 Jul 24 [Epub ahead of print]

Xueqian Xie[12] / Martin J. Willemink[3] / Yingru Zhao[12] / Pim A. de Jong[3]
Peter MA van Ooijen[12] / Matthijs Oudkerk[2]
Marcel JW Greuter[1] / Rozemarijn Vliegenthart[12]

University of Groningen, University Medical Center Groningen,
[1]Department of Radiology, [2]Center for Medical Imaging – North East Netherlands;
[3]Department of Radiology, University Medical Center Utrecht

Abstract

Objective

To assess inter- and intra-scanner variability in volumetry of solid pulmonary nodules in an anthropomorphic thoracic phantom, using low-dose computed tomography (CT).

Materials and Methods

Five spherical solid artificial nodules (diameters 3, 5, 8, 10 and 12 mm; CT density +100 HU) were randomly placed inside an anthropomorphic thoracic phantom, in different combinations. The phantom was examined on two 64-row multi-detector CT (64-MDCT) systems from different vendors with a low-dose protocol. Each CT examination was performed three times. The CT examinations were evaluated twice by independent blinded observers. Nodule volume was semi-automatically measured by dedicated software. Inter-scanner variability was evaluated by Bland-Altman analysis, and expressed as 95% confidence interval (CI) of relative differences. Intra-scanner variability was expressed as 95% CI of relative variation from the mean.

Results

No significant difference in CT-derived volume was found between CT-A and CT-B, except for the 3 mm nodules ($p < 0.05$). The 95% CI of inter-scanner variability was within ±41.6%, ±18.2% and ±4.9%, for 3, 5 and ≥ 8 mm nodules, respectively. The 95% CI of intra-scanner variability was within ±28.6%, ±13.4% and ±2.6%, for 3, 5 and ≥ 8 mm nodules, respectively.

Conclusions

Different 64-MDCT scanners in low-dose setting yield good agreement in volumetry of artificial pulmonary nodules between 5 and 12 mm diameter. Inter- and intra-scanner variability decreases at larger nodule size, to a maximum of 4.9% for ≥ 8 mm nodules.

Introduction

Lung cancer is worldwide the primary cancer in men and the second most common cancer in women, causing 18% of the total number of deaths [1]. Many lung cancers are found at a relatively late stage, resulting in a 5-year survival of only 15% or less [2]. Low-dose computed tomography (CT) is a promising screening method for early detection of lung cancer [3-7]. First results indicate that CT lung cancer screening can reduce lung cancer-specific mortality [8].

In lung cancer screening, treatment decisions usually depend on pulmonary nodule size for the nodules at first detection, and on growth rate at follow-up [4]. Therefore, it is essential to assess nodule size and growth rate accurately and reproducibly [9,10]. Variability has been found in CT-derived nodule size assessment [11, 12]. In view of the current practice that patients frequently undergo follow-up examinations, sometimes not on the same scanner, reliable inter- and intra-scanner reproducibility of nodule volumetry is important.

However, previous studies reported inconsistent results regarding the reproducibility of nodule volumetry. Some in-vitro studies have been performed in which artificial nodules were placed at known locations in a thoracic phantom without pulmonary vessels [13-15]. Some of these studies were based on older generation CT scanners [13,16]. These studies generally showed a small margin of variability in nodule volumetry for software from different vendors. On the other hand, in-vivo studies have shown that variability can be considerable, with variability up to 25% for 15 to 500 mm^3 nodules [11,17-19]. A study to investigate inter- and intra-scanner variability under optimally controlled conditions yet resembling human lungs, using a more realistic phantom, has not been performed. Nowadays, most commonly 64-row multi-detector CT (64-MDCT) scanners are utilized, also in lung cancer screening. The variability of nodule volumetry of these scanners impacts nodule management, for example the interval of repeated CT scanning. As an extension to our recent study on observer detection and accuracy of manual and semi-automated volumetry [10], the focus of this study is on reproducibility between and within 64-MDCT systems. We assessed the inter- and intra-scanner variability of pulmonary nodule volumetry on low-dose 64-MDCT, using randomly placed solid nodules in an anthropomorphic thoracic phantom with a background of pulmonary vasculature.

Materials and Methods

Phantom

This study was performed using an anthropomorphic thoracic phantom (Lungman, Kyoto Kagaku, Tokyo, Japan) with an artificial thoracic wall, heart, mediastinum, diaphragm and lungs with pulmonary vasculature (Figure 1). This phantom simulates an accurate life-size anatomical model of a male thorax. Soft tissue substitute materials were made of polyurethane resin composites. Synthetic bones were made of epoxy resin, with x-ray absorption rates close to those of human tissue. The space between the pulmonary vessels in the thoracic cavity contained air.

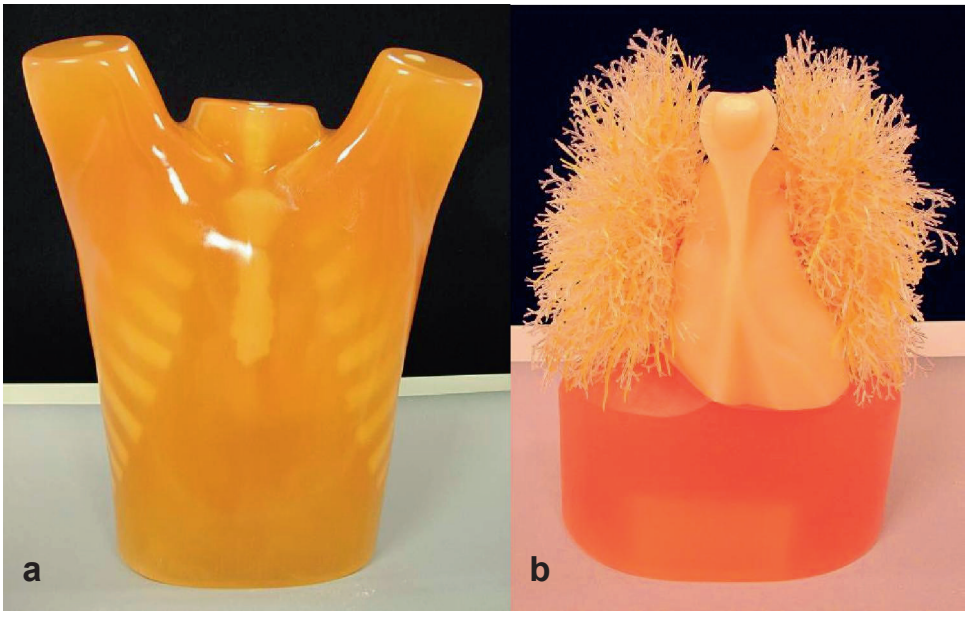


Figure 1 External view (**a**) and internal artificial heart and lungs (**b**) of the anthropomorphic thoracic phantom.

Moreover, we used five commercially available artificial spherical pulmonary nodules (Kyoto Kagaku, Tokyo, Japan) with a smooth surface in five diameters (3, 5, 8, 10 and 12 mm, corresponding to a volume of 14, 65, 268, 524 and 905 mm^3), made of polyurethane resin (Figure 2). CT density was +100 Hounsfield units (HU) at 120 kV.

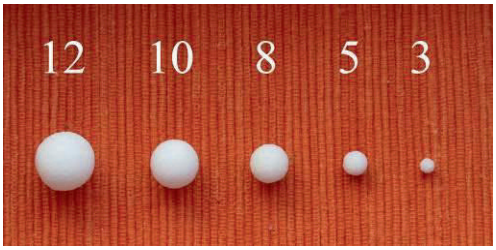


Figure 2 Artificial spherical pulmonary nodules with five different diameters (from left to right, 12, 10, 8, 5 and 3 mm, corresponding to volumes of 905, 524, 268, 65 and 14 mm^3).

CT imaging

Two 64-MDCT systems (CT-A: Sensation 64, Siemens, Forchheim, Germany; CT-B: Brilliance 64, Philips, Best, The Netherland) were utilized. The low-dose CT acquisition protocol in the Dutch-Belgian Randomized Lung Cancer Screening Trial (NELSON) was applied [4]. The protocol for CT-A was: spiral acquisition at 120 kV, 20 mAs, rotation time 0.5 s, pitch 1.5 and slice collimation 2 × 32 × 0.6 mm, field of view 300 mm; images were reconstructed at a slice thickness of 1 mm, a slice increment of 0.7 mm, using a medium-smooth B30f image reconstruction kernel. The protocol for CT-B was: spiral acquisition at 120 kV, 20 mAs, rotation time 0.5 s, pitch 1.39 and slice collimation 64 × 0.625 mm, field of view 300 mm; images were reconstructed at a slice thickness of 1 mm, a slice increment of 0.7 mm, with a medium-smooth B kernel. Both scanners were routinely calibrated.

Artificial nodules were randomly positioned in the pulmonary vessels of the phantom. All nodules were firmly held in the vessels without using glue. To avoid the influence from thoracic walls on nodule segmentation, all nodules were attached to pulmonary vessels, none of the nodules were attached to pleura or positioned subpleurally. After randomly positioning a predetermined set of one to three nodules in the artificial lungs, CT image acquisition was performed. Thirteen different combinations of nodules were set up so that each nodule was examined in total five times. The CT examination was repeated three times for each nodule configuration, with a small translocation (approximately 5 to 10 mm) and rotation (approximately 5 to 10 degrees) of the phantom in between each examination to simulate participant movement. The thoracic phantom was also examined once without pulmonary nodules, serving as a control examination to exclude false positive findings. Thus, per CT system 40 examinations were performed, including 39 test examinations and one control examination.

Therefore, each nodule was examined 30 times (five different locations, three repeating examinations and two CT systems). Furthermore, all nodules were examined in air, on the CT table, to confirm the visibility of the artificial pulmonary nodules on low-dose CT without the artificial thoracic background.

Quantitative image analysis

The reconstructed data were evaluated on a workstation (Leonardo, Siemens, Forchheim, Germany). Three observers participated in image analysis, including two radiologists specialized in thoracic diagnostic imaging with eight years experiences (observer-A and observer-B), and one resident with two years experiences (observer-C). All the examinations were assessed twice separated by at least one month by all three observers. The data were presented in the same order for the double reading. Observer-A evaluated the examinations of CT-A and CT-B. Observer-B evaluated the examinations of CT-A. Observer-C evaluated the examinations of CT-B. The three observers were blinded to information about the presence, properties and location of the artificial pulmonary nodules. Since each nodule was scanned 15 times on each CT system and measured twice, in total at most 30 paired evaluations for each nodule could be included for inter-scanner analysis. Also 30 evaluations for each nodule could be included for intra-scanner analysis.

The observers were asked to report whether there were pulmonary nodules present or not. If a potential nodule was observed, the images were compared to the images of the control CT examination without a nodule to confirm that it was not a false positive finding caused by pulmonary background structures. If comparison with the control CT confirmed the presence of the pulmonary nodule, a dedicated commercial software tool (LungCARE, Siemens, Forchheim, Germany) was utilized to semi-automatically measure the volume of the detected nodules. The accuracy of nodule volumetry of this software was shown to be very high [20].

In addition, image noise was assessed by measuring the standard deviation of mean HU in a circular region of interest (ROI) with a radius of 10 mm. The ROIs were placed in artificial left ventricle, avoiding image artifact [21]. Ten examinations per CT scanner were randomly selected. Three ROIs were measured in each examination.

Statistics

The normality of data was assessed by the Kolmogorov-Smirnov test. Influential effects on nodule volumetry were investigated using univariate analysis of a general linear model. The dependent variable was CT-derived volume. The fixed factors were CT system (CT-A and CT-B) and observer (Observer-A, Observer-B and Observer-C). Differences in CT-derived volumes between the two CT systems were evaluated by a paired-samples t-test.

Inter-scanner reliability of CT-derived volume was expressed as an intraclass correlation coefficient (ICC). An ICC value larger than 0.90 was considered as high agreement, in the range of 0.75 to 0.90 as moderate, and smaller than 0.75 as low [22]. Inter-scanner variability of CT-derived volume was explored using Bland-Altman plots. Relative difference was calculated as $(a - b) / m_1 \times 100\%$, in which, a and b were CT-derived volumes on CT-A and CT-B, respectively. m_1 was the mean of a and b. The 95% confidence interval (CI) of

relative difference was calculated as a mean relative difference ± 1.96 standard deviation (SD).

Since multiple image acquisitions and measurements were performed for an individual nodule, intra-scanner variability was expressed as a variation coefficient, by calculating SD/m_2, where m_2 was the mean of 30 evaluations of each nodule [16]. Relative variation from the mean was calculated. The 95% CI of relative variation was calculated as (± 1.96 SD) / m_2 × 100% [23]. A 95% CI smaller than 25% was considered as good agreement for inter- and intra-scanner variability [17].

Results were given as mean ± SD. A $p < 0.05$ was considered as statistically significant. All statistical analyses were performed using SPSS version 20 (IBM, New York, US).

Results

All the nodules were detected and measured by the two observers. No false positive nodules were found. Representative CT images of the anthropomorphic thoracic phantom are shown in Figure 3. CT-derived volume for each artificial nodule conformed to a normal distribution (test for normality $p > 0.05$).

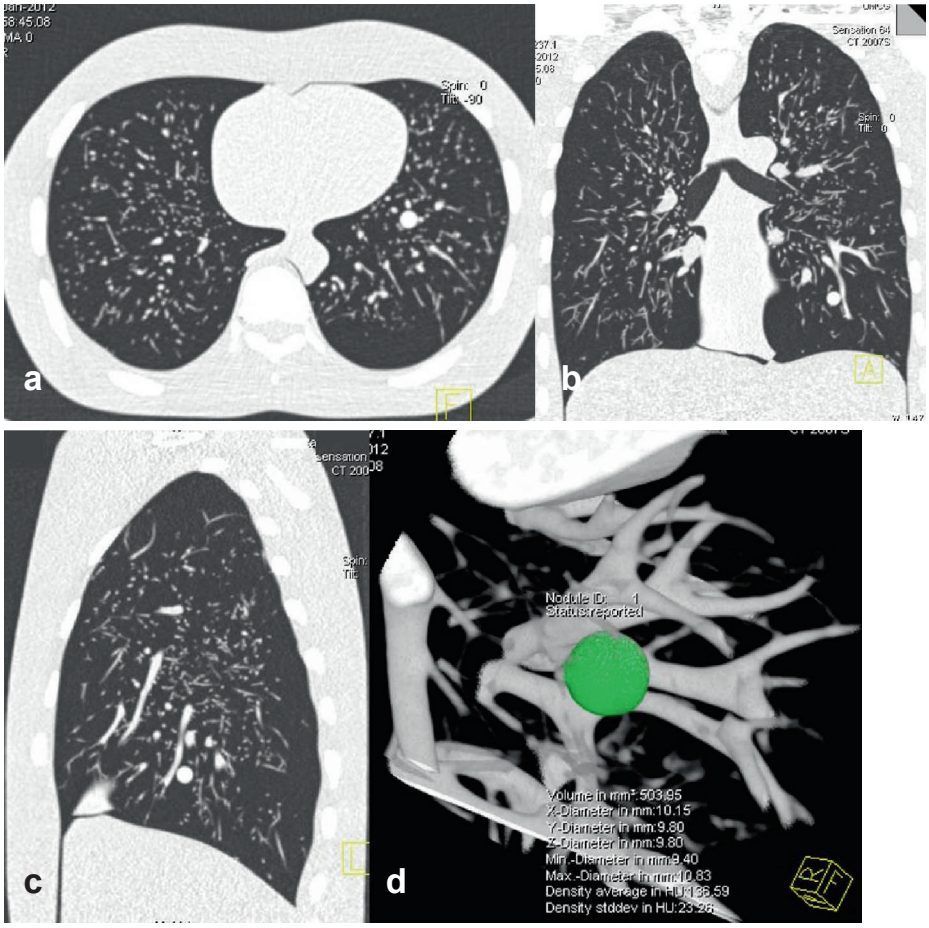


Figure 3 Axial (**a**), coronal (**b**) and sagital image (**c**) of the anthropomorphic thoracic phantom with a semi-automatically measured solid pulmonary nodule (**d**).

Table 1 Inter-scanner variability of CT-derived volume for artificial spherical solid pulmonary nodules

Nodule characteristics		Relative difference	
Diameter, mm	Volume, mm^3	Mean, %	95% CI, %
3	14	16.7	-19.3, 41.6
5	65	7.2	-5.6, 18.2
8	268	0.9	-2.3, 4.0
10	524	2.2	-0.5, 4.9
12	905	1.3	-1.1, 3.7

Relative difference was calculated as (volume on CT-A – volume on CT-B) / (mean volume of CT-A and CT-B) × 100%. 95% CI was calculated as mean ± 1.96 SD.

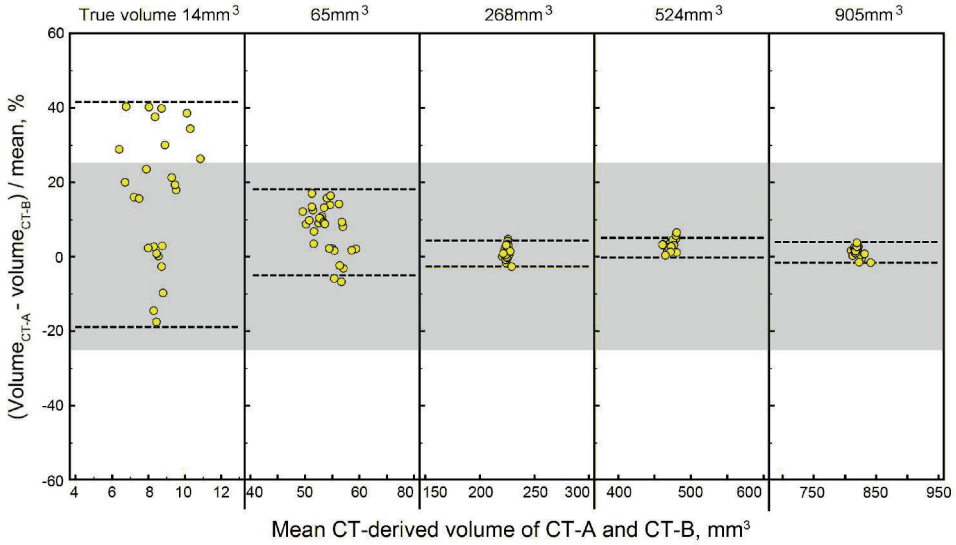


Figure 4 Bland-Altman plots of the relative inter-scanner difference in semi-automated volumetry, for solid (+100 HU) nodules in five sizes (3, 5, 8, 10 and 12 mm in diameter, corresponding to volumes of 14, 65, 268, 524 and 905 mm^3) as measured on two CT systems of two different vendors, CT-A and CT-B. The dotted lines indicate the 95% confidence interval. The grey zone indicates a relative difference of ±25%, which is commonly used to exclude systematic errors, thus to exclude a growing nodule on subsequent CT scans.

Volumetry

Univariate analysis showed that neither CT system nor observer was a significant factor influencing volumetry ($p > 0.05$). No significant difference in CT-derived volume was found between CT-A and CT-B ($p > 0.05$), except for the smallest 3 mm nodules ($p < 0.05$). Semi-automated volumetry demonstrated underestimation from the actual volume. Overall, CT-derived volume was underestimated by 9.1 ± 7.9% ($p < 0.001$). The noise level in CT-A was 21.8 ± 1.3, significantly lower than 24.0 ± 1.3 in CT-B ($p < 0.01$).

Inter- and intra-scanner variability

Inter-scanner reliability of CT-derived volume was high (ICC = 1.000, $p < 0.001$). Inter-scanner variability of CT-derived volume decreased at larger nodule size (Table 1, Figure 4). Inter-scanner variability was low for nodules ≥ 8 mm in diameter, in which the maxima of the 95% CI was 4.9%.

Table 2 Intra-scanner variability of CT-derived volume for artificial spherical solid pulmonary nodules

Nodule characteristics		CT-A		CT-B	
Diameter, mm	Volume, mm^3	Variation coefficient,%	95% CI of relative variation, %	Variation coefficient,%	95% CI of relative variation, %
3	14	14.6	-28.6, 28.6	14.6	-28.6, 28.6
5	65	3.9	-7.6, 7.6	6.8	-13.4, 13.4
8	268	1.1	-2.2, 2.2	1.1	-2.2, 2.2
10	524	1.3	-2.6, 2.6	0.7	-1.4, 1.4
12	905	0.6	-1.3, 1.3	1.1	-2.2, 2.2

Variation coefficient was calculated as SD / mean × 100%. 95% CI of relative variations was calculated as (±1.96 SD) / mean × 100%.

Intra-scanner variability also decreased at larger nodule size (Table 2, Figure 5). Intra-scanner variability of CT-derived volume was also low for nodules ≥ 8 mm in diameter, in which the maxima of the 95% CI was 2.6%.

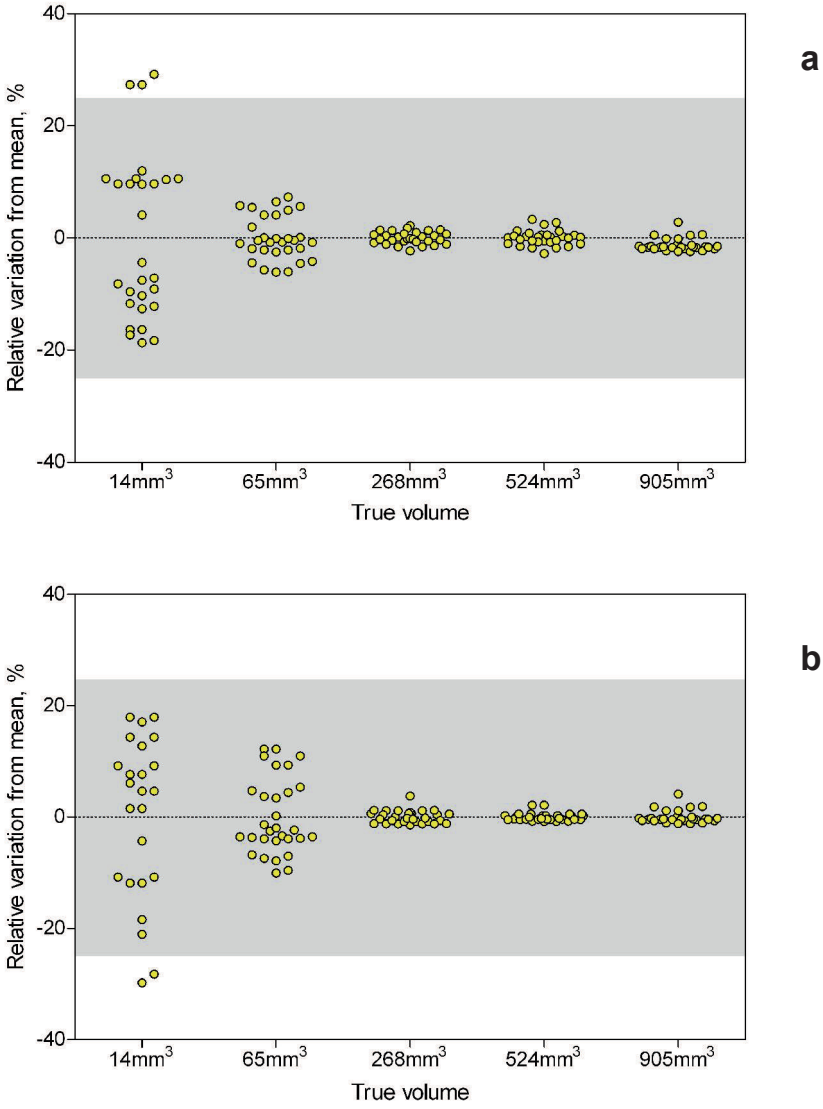


Figure 5 Grouped plots of the relative intra-scanner variation in semi-automated volumetry, for solid (+100 HU) nodules in five sizes (3, 5, 8, 10 and 12 mm in diameter, corresponding to volumes of 14, 65, 268, 524 and 905 mm^3) on CT-A (**a**) and CT-B (**b**). The grey zone indicates a relative difference of ±25%, which is commonly used to exclude systematic errors, thus to exclude a growing nodule on subsequent CT scans.

Discussion

In this anthropomorphic phantom study investigating inter- and intra-scanner variability, and inter-measurement variability of CT-derived nodule volume, variability of CT-derived volume decreased at increasing nodule size, to a maximum of 4.9% for ≥ 8 mm nodules. Good inter- and intra-scanner agreement was found for nodules ≥ 5 mm in diameter, but not for 3 mm nodules.

One-dimensional measurement (diameter) of a target lesion is commonly utilized [24]. Yet accuracy and reproducibility of diameter measurement to assess pulmonary nodule size are low [10]. Also, discrepancies in manual assessment of lesion size can significantly influence therapeutic decisions [25]. On the other side, semi-automated volumetry using dedicated software shows better reproducibility than manual assessment of nodule size [26, 27]. However, even semi-automated volumetry suffers from some variability [11, 12]. It is well known that reproducibility of CT-derived volume of pulmonary nodules is influenced by a number of factors, such as CT equipment, nodule characteristics and pulmonary surroundings [9, 11]. In view of current practice of lung cancer screening that uses multiple CT systems, management of systematic errors is essential to reach a reliable result.

Variability of CT-derived volume by semi-automated method decreased at larger nodule size. In previous in-vivo studies, pulmonary nodules were commonly analyzed as a whole, without assessing variability by nodule size categories [17, 18]. Also, inconsistent results were reported about dependency of volumetry variability on nodule size. Rampinelli et al. found that volumetry for larger nodules was more reproducible, based on 83 nodules with a mean volume of 220 ± 241 mm^3 [19], yet Marchiano et al. observed that variability was not affected by nodule volume based on 233 nodules of 99 ± 127 mm^3 [28]. In those in-vivo studies, different volume ranges of included nodules have likely caused these inconsistent results. We performed the current study based on five discrete and well-defined sizes of nodules from 14 to 905 mm^3, which is the common nodule volume range in lung cancer screening [4], and found that variability does depend on nodule size.

Currently, a threshold of 25% is commonly utilized as the minimal volume change for growing solid nodules [3]. That in-vivo result was based on inter-scan variability observed in one CT system for nodules 15 to 500 mm^3 [17]. Our results were based on two CT systems, which is closer to the practice of lung cancer screening with multiple CT systems involved. In our study, the maxima of 95% CI of variability was less than 18% for ≥ 5 mm (65 mm^3) solid nodules, and less than 5% for ≥ 8 mm (268 mm^3) nodules. This suggests that the threshold of 25% can be used for solid nodules of ≥ 5 mm and is a conservative threshold. The threshold of 25% could potentially be decreased for larger pulmonary nodules of ≥ 8 mm. Because large variability was found for 3 mm nodules, the threshold of 25% is not

likely to be usable for 3 mm nodules. However, the likelihood of < 5 mm nodules to be malignant is very low [29].

In lung cancer screening, images are often read twice, commonly by different paired observers. E.g., in the NELSON trial, the CT examination of participants was assessed first at the local site, and then at the central site. Two radiologists with more than 6 years experience performed the evaluation at the central site, and 13 radiologists or residents with variable experience at the local sites [30]. To adhere to this screening practice of double reading, we chose two observers (local readers) for the first evaluation and one observer (central reader) for the second evaluation.

A significant volumetry difference was observed between CT-A and CT-B in 3 mm nodule, but not in the larger nodules. This was caused by different properties between the scanners from both vendors, i.e., image acquisition and reconstruction parameters. Because volumetry analysis of smaller nodules is more affected by noise, we quantified the image noise in the images from both CT scanners. Higher image noise degrades nodule segmentation, thus leading to underestimated CT-derived volume [31]. As we observed, CT-B has higher noise level than CT-A, this could very well explain the observed lower CT-derived volume in CT-B than in CT-A for the 3 mm nodule.

Semi-automated volumetry systematically underestimated the nodule volume when using a low-dose CT acquisition protocol. This finding is in concordance with previous studies based on the same phantom, in three different CT scanners [10, 32]. It is well know that accuracy of nodule volumetry depends on many factors, such as CT acquisition and reconstruction algorithm. De Jong et al. reported that low-dose CT yielded lower volumes than normal-dose CT [33]. It is likely that our low-dose setting contributes to the systematic volume underestimation. However, given the systematic nature of this error, this is less important to evaluate growth rate in lung cancer screening.

Clinical implications

Changes in nodule size that exceed measurement variability are important to determine actual growth, which can impact treatment decision and estimation of therapy outcome [34, 35]. In lung cancer screening trials, volumetry and diameter methods to assess pulmonary nodule size and growth have been applied. Semi-automated volumetry was utilized in the NELSON trial [36]. Manual diameter measurement was utilized in some earlier launched trials, such as the National Lung Screening Trial [8]. Since determination of nodule growth depends on nodule size, a volumetry method with less variability can potentially provide more accurate and more timely treatment decisions and a better outcome estimation. We found that semi-automated volumetry showed good inter- and intra-scanner variability. Thus, the semi-automated method is reliable for pulmonary nodule assessment in lung cancer screening. Moreover, no difference was found between two CT

systems from different vendors for ≥ 5 mm nodules. As the follow-up of screened participants can last for an extensive period, different CT systems from different vendors may be used. A direct comparison of nodule volumes obtained from different CT systems seems valid, at least for the CT systems investigated in this study.

To determine a growing nodule, inter- and intra-scanner variability should be optimized to minimize systematic errors. This allows for the determination of a cut-off value to assess with confidence whether nodule volumes have actually increased. For an indeterminate nodule, lung cancer screening trials commonly use a follow-up CT of three to fourth months to evaluate nodule growth, based on a minimum increase of 25% [6]. The threshold of 25% can likely be lowered for nodules of ≥8 mm in diameter, when using semi-automated volumetry and 64-MDCT scanning, since the variability was only 4.9% in this study. If based on a minimum increase of 5% to determine growth, the follow-up interval can potentially be reduced to three to four weeks. This can also limit the time of participant anxiety.

Limitations

Firstly, although the anthropomorphic phantom more closely resembles the human thorax than previously published phantoms, degenerative changes in lung tissue were not present in the phantom. Fibrosis, emphysema and consolidations can influence the nodule volumetry. Additional patient factors were not present either, such as breathing variability, variable habitus and patient size. A phantom with these variable characteristics is difficult to make. Our phantom study represents the variability of nodule volumetry in a relatively optimal environment, although the variability might be higher in vivo. Also, only spherical nodules of five discrete sizes were used. Although pulmonary nodules of these sizes are common findings in lung cancer screening [4], nodule shape is more often irregular and nonspherical in-vivo, with a continuous size range. A study extension would therefore be variability assessment for irregular nodules.

Secondly, we only evaluated one CT acquisition protocol. That protocol has been widely applied in lung cancer screening [3].However, variability of nodule volumetry depends on acquisition and algorithm in individual settings [7, 35]. When discussing the results, we linked to the studies based on the same CT acquisition protocol [5, 10, 11, 17]. Thus, our results are comparable to those previous data. A further extension would be the evaluation of different CT acquisition protocols.

Conclusions

Different 64-MDCT scanners in low-dose setting yield similar volumetry of artificial pulmonary nodules between 5 and 12 mm diameter. Inter- and intra-scanner variability, and inter-measurement variability of CT-derived volumes of spherical solid pulmonary nodules decreases at increasing nodule size. In lung cancer screening, the commonly accepted cut-off of 25% to determine nodule growth has the potential to be reduced for nodules of ≥ 8 mm in diameter. This phantom study offers potential to reduce the interval for repeated CT scans in lung cancer screening.

References

1. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer. 2010;127:2893-917.
2. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. CA Cancer J Clin. 2012;62:10-29.
3. Nair A, Hansell DM. European and North American lung cancer screening experience and implications for pulmonary nodule management. Eur Radiol. 2011;21:2445-54.
4. van Klaveren RJ, Oudkerk M, Prokop M, Scholten ET, Nackaerts K, Vernhout R, et al. Management of lung nodules detected by volume CT scanning. N Engl J Med. 2009;361:2221-9.
5. Xu DM, van Klaveren RJ, de Bock GH, Leusveld A, Zhao Y, Wang Y, et al. Limited value of shape, margin and CT density in the discrimination between benign and malignant screen detected solid pulmonary nodules of the NELSON trial. Eur J Radiol. 2008;68:347-52.
6. Yingru Z, de Bock GH, Vliegenthart R, van Klaveren RJ, Ying W, Bogoni L, et al. Performance of computer-aided detection of pulmonary nodules in low-dose CT: comparison with double reading by nodule volume. Eur Radiol. 2012;22:2076-84.
7. Kakinuma R, Ashizawa K, Kobayashi T, Fukushima A, Hayashi H, Kondo T, et al. Comparison of sensitivity of lung nodule detection between radiologists and technologists on low-dose CT lung cancer screening images. Br J Radiol. 2012;85:E603-E08.
8. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med. 2011;365:395-409.
9. Gavrielides MA, Kinnard LM, Myers KJ, Petrick N. Noncalcified lung nodules: volumetric assessment with thoracic CT. Radiology. 2009;251:26-37.

10. Xie X, Zhao Y, Snijder R, van Ooijen P, Oudkerk M, de Bock G, et al. Sensitivity and accuracy of volumetry of pulmonary nodules on low-dose 16- and 64-row multidetector CT: an anthropomorphic phantom study. Eur Radiol 2013;23:139-47.
11. Wang Y, van Klaveren RJ, van der Zaag-Loonen HJ, de Bock GH, Gietema HA, Xu DM, et al. Effect of nodule characteristics on variability of semiautomated volume measurements in pulmonary nodules detected in a lung cancer screening program. Radiology. 2008;248:625-31.
12. Gietema HA, Wang Y, Xu D, van Klaveren RJ, deKoning H, Scholten E, et al. Pulmonary nodules detected at lung cancer screening: Interobserver variability of semiautomated volume measurements. Radiology. 2006;241:251-57.
13. Das M, Ley-Zaporozhan J, Gietema HA, Czech A, Muehlenbruch G, Mahnken AH, et al. Accuracy of automated volumetry of pulmonary nodules across different multislice CT scanners. Eur Radiol. 2007;17:1979-84.
14. Marten K, Dullin C, Machann W, Schmidt JS, Das M, Hermann KP, et al. Comparison of flat-panel-detector-based CT and multidetector-row CT in automated volumetry of pulmonary nodules using an anthropomorphic chest phantom. Br J Radiol. 2009;82:716-23.
15. Goodsitt MM, Chan HP, Way TW, Schipper MJ, Larson SC, Christodoulou EG. Quantitative CT of lung nodules: Dependence of calibration on patient body size, anatomic region, and calibration nodule size for single- and dual-energy techniques. Med Phys. 2009;36:3107-21.
16. Bolte H, Riedel C, Jahnke T, Inan N, Freitag S, Kohl G, et al. Reproducibility of computer-aided volumetry of artificial small pulmonary nodules in ex vivo porcine lungs. Invest Radiol. 2006;41:28-35.
17. Gietema HA, Schaefer-Prokop CM, Mali WPTM, Groenewegen G, Prokop M. Pulmonary nodules: Interscan variability of semiautomated volume measurements with multisection CT-influence of inspiration level, nodule size, and segmentation performance. Radiology. 2007;245:888-94.
18. Wormanns D, Kohl G, Klotz E, Marheine A, Beyer F, Heindel W, et al. Volumetric measurements of pulmonary nodules at multi-row detector CT: in vivo reproducibility. Eur Radiol. 2004;14:86-92.
19. Rampinelli C, De Fiori E, Raimondi S, Veronesi G, Bellomi M. In vivo repeatability of automated volume calculations of small pulmonary nodules with CT. AJR Am J Roentgenol. 2009;192:1657-61.
20. Bolte H, Riedel C, Muller-Hulsbeck S, Freitag-Wolf S, Kohl G, Drews T, et al. Precision of computer-aided volumetry of artificial small solid pulmonary nodules in ex vivo porcine lungs. Br J Radiol. 2007;80:414-21.

21. Christe A, Charimo-Torrente J, Roychoudhury K, Vock P, Roos JE. Accuracy of low-dose computed tomography (CT) for detecting and characterizing the most common CT-patterns of pulmonary disease. Eur J Radiol. 2013;82:E142-E50.
22. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33:159-74.
23. Paarmann H, Groesdonk HV, Sedemund-Adib B, Hanke T, Heinze H, Heringlake M, et al. Lack of agreement between pulmonary arterial thermodilution cardiac output and the pressure recording analytical method in postoperative cardiac surgery patients. Br J Anaesth. 2011;106:475-81.
24. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). Eur J Cancer. 2009;45:228-47.
25. Pauls S, Ktirschner C, Dharaiya E, Muche R, Schmidt SA, Krueger S, et al. Comparison of manual and automated size measurements of lung metastases on MDCT images: Potential influence on therapeutic decisions. Eur J Radiol. 2008;66:19-26.
26. Hein PA, Romano VC, Rogalla P, Klessen C, Lembcke A, Dicken V, et al. Linear and volume measurements of pulmonary nodules at different CT dose levels - intrascan and interscan analysis. RoFo. 2009;181:24-31.
27. Wiemker R, Rogalla P, Blaffert T, Sifri D, Hay O, Shah E, et al. Aspects of computer-aided detection (CAD) and volumetry of pulmonary nodules using multislice CT. Br J Radiol. 2005;78:S46-S56.
28. Marchiano A, Calabro E, Civelli E, Di Tolla G, Frigerio LF, Morosi C, et al. Pulmonary nodules: volume repeatability at multidetector CT lung cancer screening. Radiology. 2009;251:919-25.
29. Xu DM, van der Zaag-Loonen HJ, Oudkerk M, Wang Y, Vliegenthart R, Scholten ET, et al. Smooth or attached solid indeterminate nodules detected at baseline CT screening in the NELSON study: cancer risk during 1 year of follow-up. Radiology. 2009;250:264-72.
30. Wang Y, van Klaveren RJ, de Bock GH, Zhao Y, Vernhout R, Leusveld A, et al. No benefit for consensus double reading at baseline screening for lung cancer with the use of semiautomated volumetry software. Radiology. 2012;262:320-6.
31. Chen B, Barnhart H, Richard S, Colsher J, Amurao M, Samei E. Quantitative CT: technique dependence of volume estimation on pulmonary nodules. Phys Med Biol. 2012;57:1335-48.
32. Willemink MJ, Leiner T, Budde RPJ, de Kort FPL, Vliegenthart R, van Ooijen PMA, et al. Systematic error in lung nodule volumetry: effect of iterative reconstruction versus filtered back projection at different CT parameters. AJR Am J Roentgenol. 2012;199:1241-46.

33. de Jong PA, Leiner T, Lammers JW, Gietema HA. Can low-dose unenhanced chest CT be used for follow-up of lung nodules? AJR Am J Roentgenol. 2012;199:777-80.
34. Marten K, Auer F, Schmidt S, Rummeny EJ, Engelke C. Automated CT volumetry of pulmonary metastases: the effect of a reduced growth threshold and target lesion number on the reliability of therapy response assessment using RECIST criteria. Eur Radiol. 2007;17:2561-71.
35. Macpherson RE, Higgins GS, Murchison JT, Wallace WAH, Price A, Gaffney S, et al. Non-small-cell lung cancer dimensions: CT-pathological correlation and interobserver variation. Br J Radiol. 2009;82:421-25.
36. Ru Zhao Y, Xie X, de Koning HJ, Mali WP, Vliegenthart R, Oudkerk M. NELSON lung cancer screening study. Cancer imaging. 2011;11:S79-84.

Chapter 6

Validation and Prognosis of Coronary Artery Calcium Scoring in Nontriggered Thoracic Computed Tomography: Systematic Review and Meta-Analysis

Circulation Cardiovascular Imaging. 2013; 6: 514-521

Xueqian Xie[12] / Yingru Zhao[12] / Geertruida H. de Bock[3] / Pim A. de Jong[4]
Willem P. Mali[4] / Matthijs Oudkerk[2] / Rozemarijn Vliegenthart[12]

University of Groningen, University Medical Center Groningen, [1]Department of Radiology, [2]Center for Medical Imaging–North East Netherlands, [3]Department of Epidemiology; [4]Department of Radiology, University Medical Center Utrecht

Abstract

Background

Coronary calcium score (CS), traditionally based on electrocardiography (ECG)-triggered computed tomography (CT), predicts cardiovascular risk. Nowadays, nontriggered thoracic CT is extensively utilized, such as in lung cancer screening. The study-purpose was to determine the correlation in CS between nontriggered and ECG-triggered CT, and to evaluate the prognostic performance of the CS derived from nontriggered CT.

Methods and Results

PubMed, Embase and Web of Knowledge were searched until November, 2012. Two reviewers independently screened 2,120 records to identify studies reporting the CS in nontriggered CT, and extracted information. Study quality was evaluated by standardized assessment tools. Cohen's κ was extracted for agreement of CS categories between nontriggered and ECG-triggered CT (validation). Hazard ratio was extracted for prognostic performance. Five studies comprising 1,316 individuals were included regarding validation. Five studies comprising 34,028 cardiac asymptomatic individuals, mainly from lung cancer screening trials, were included regarding prognosis. All studies were of high quality. Meta-analysis could only be performed for validation studies, as studies on prognostic performance were highly heterogeneous. Pooled Cohen's κ for agreement between the two techniques was 0.89 (95%CI: 0.83 to 0.95) for increasing CS categories. Increasing CS categories were associated with increasing risk of cardiovascular death or events. Nontriggered CT yielded false-negative CS in 8.8% of individuals, and underestimated high CS in 19.1%.

Conclusions

Our analysis shows the prognostic value and potential role of nontriggered assessment of coronary calcium, but it does not suggest that ECG-triggered CT should be replaced by nontriggered exams.

Introduction

The amount of coronary artery calcium, based on computed tomography (CT) and traditionally expressed as calcium score (CS) according to Agatston [1], is a strong predictor of cardiovascular events [2-5]. Calcium scoring has been found to improve cardiovascular risk stratification beyond cardiovascular risk factors [4,6]. Due to the irregular and periodic movements of coronary arteries, electrocardiography (ECG)-triggered cardiac acquisition techniques are applied in CT to minimize motion artifacts and optimize calcium scoring [3].

Compared to ECG-triggered CT, nontriggered CT is extensively utilized. In 2007, 13.6 million nontriggered thoracic CT examinations were performed in the United States, in contrast to 0.7 million ECG-triggered CT examinations for calcium scoring [7]. Recent trial results have increased the interest in lung cancer screening by thoracic CT [8]. Thus, the number of nontriggered examinations will likely further increase. Age and smoking, the current selection criteria for lung cancer screening, are also correlated with coronary calcification and coronary heart disease [9]. In lung cancer screening, coronary calcification is a frequent finding [10]. If nontriggered CT can be used for calcium scoring, to stratify individuals in categories of cardiovascular risk and to identify those at high cardiovascular risk, there may be a substantial unused primary prevention potential [11]. Also, deriving the CS from the same examination as used in lung cancer screening may positively impact the cost-effectiveness of screening.

Because motion of coronary arteries influences calcium scoring [12], the utilization of coronary calcium scoring in nontriggered CT is still being debated [13]. With the increasing interest in lung cancer screening, this is an optimal moment to investigate the potential utilization of nontriggered CT for calcium scoring. However, compared with the extensive publications in ECG-triggered cardiac CT, the literature on calcium scoring in nontriggered thoracic CT is relatively limited. Therefore, we conducted a systematic review and meta-analysis to investigate the validity and prognostic value of calcium scoring derived from nontriggered thoracic CT.

Methods

This study was conducted according to Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) [14].

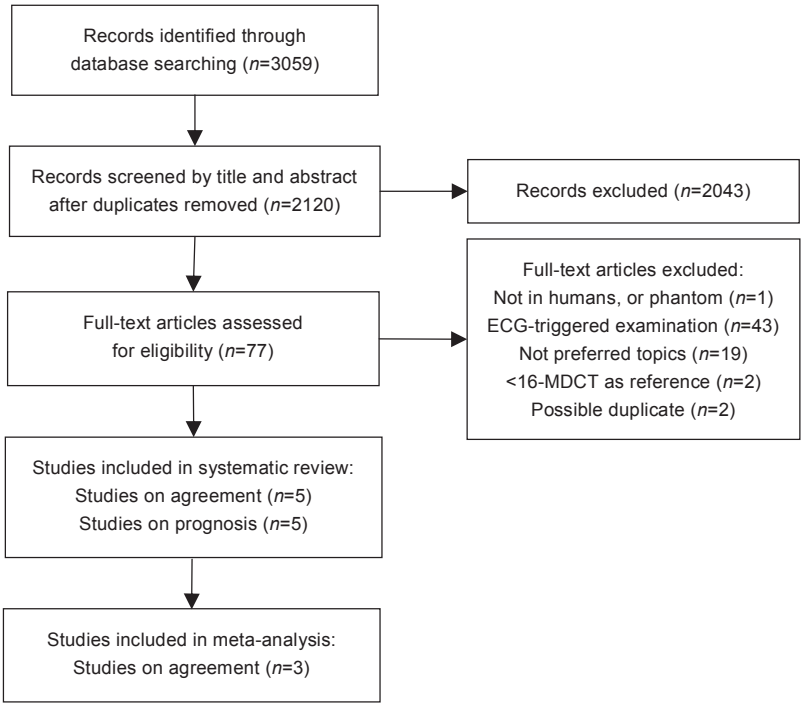


Figure 1 Flowchart of literature review and selection.

Information sources and search

We searched PubMed, Embase, and Web of Knowledge until November, 2012, using terms related to computed tomography and nontriggered thoracic examination (ungated, untriggered, nongated, nontriggered, thorax, chest, thoracic, pulmonary, etc.) and coronary and calcium (calcium or calcification, etc.) without language restrictions (Supplemental Table A). Unpublished studies were not included.

Study selection

Two reviewers (XQ.X. and YR.Z.) with at least eight years of experience in thoracic and cardiovascular radiology participated in literature selection. Each record was evaluated independently. Disagreement in literature selection was resolved by consensus. Studies were included in the systematic review when they: 1) evaluated cardiac asymptomatic adult humans, or phantoms; 2) analyzed one of the following topics regarding calcium scoring in nontriggered CT: agreement between nontriggered and ECG-triggered CT, or prognostic performance to predict death or events; 3) used at least 16-row MDCT as ECG-triggered examination when ECG-triggered CT was used as reference examination. Sixteen-row MDCT was used as minimum CT generation because previous research showed higher ac-

curacy and reproducibility in calcium scoring for 16-row MDCT compared to earlier generation CT systems [15].

Articles were excluded when they: 1) were reviews, abstracts, case reports or letters; 2) investigated participants with confounding factors, e.g., pacemaker or defibrillator implant, and cardiac surgery. When multiple similar publications based on the same trial were identified, only the study with the largest sample size was included to avoid possible duplicate reporting.

Subsequently, meta-analysis was performed in studies on agreement between nontriggered and ECG-triggered CT, when the studies used the same calcium scoring method, i.e., continuous CS and/or four CS categories (0, 1 - 99, 100 - 399, ≥ 400). No meta-analysis could be performed of the studies on prognostic value, because of large heterogeneity in calcium quantification methods, CS categorization and outcomes.

Data collection process

A standardized data extraction form was used to collect study and participant characteristics, methodology, and main results. Two reviewers (XQ.X. and YR.Z.) collected data independently. Disagreement in data collection was resolved by consensus.

For results of studies on agreement of calcium scoring between nontriggered and ECG-triggered CT, a correlation coefficient (CC) was extracted for continuous data, and Cohen's κ and concordance percentage were extracted for categorical data. When available, the subject number with CS of > 0, < 400 and ≥ 400 was extracted for the two techniques. A CS of ≥ 400 is commonly considered as indicating high cardiovascular risk [3,5]. Thereafter, the diagnostic performance of nontriggered CT was calculated using ECG-triggered CT as reference. The percentage of false negative CS was calculated as the percentage of subjects with zero CS in nontriggered CT among subjects with CS > 0 in ECG-triggered CT. The percentage of underestimated high-risk CS was considered as the percentage of subjects with CS < 400 in nontriggered CT among subjects with CS ≥ 400 in ECG-triggered CT. The percentage of overestimated high-risk CS was calculated as the percentage of subjects with CS ≥ 400 in nontriggered CT among subjects with CS < 400 in ECG-triggered CT.

For prognostic performance of calcium scoring in nontriggered CT, hazard ratio (HR) for increasing CS categories derived from nontriggered CT to predict cardiovascular death or cardiovascular events (coronary heart disease, cerebrovascular disease, heart failure, peripheral arterial disease, aortic aneurysm, etc.) was extracted. When possible, unadjusted and adjusted HR with 95% confidence interval (CI) was extracted. Furthermore, the number of subjects with zero CS was extracted, as well as the number of subsequent cardiovascular deaths or events among these subjects.

Study quality assessment

Two reviewers (XQ.X. and YR.Z.) evaluated study quality independently on the studies included in the systematic review. Disagreement in quality assessment was resolved by consensus. Two quality assessment tools for different type of study were utilized to evaluate methodological quality and potential sources of bias, as described below.

For validation studies on agreement between nontriggered and ECG-triggered CT, the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) with 14 standard items was utilized [16]. For each study, a quality score was derived by assigning 1 point to each fulfilled item, 0.5 to an unclear item, and 0 to an unmet item, with a total possible score of 14 (Supplemental Table B).

For prognostic studies, the quality assessment criteria to evaluate reports on prognosis of CAC in American College of Cardiology Foundation / American Heart Association (ACCF / AHA) guideline with 8 standard criteria was utilized [5]. For each study, a quality score was derived by assigning 0 to 3 points to each criterion, with a total possible score of 16 (Supplemental Table C).

Synthesis of results and risk of bias

The pooling calculations of agreement between the two techniques were performed using the Hedges-Vevea random effects model and Z-test for overall effect. The pooling calculation was performed if there were at least two studies reporting the same measurement. Heterogeneity was tested using Q statistic and I^2 index. A two-sided p value for Q statistic < 0.10 or $I^2 > 50\%$ was considered to indicate heterogeneity. The random effects model was used regardless of the heterogeneity test, although results in Q statistics and I^2 index were still stated. Publication bias was evaluated with the Begg and Mazumdar rank correlation and Egger's regression test if the number of effect size in the included studies was at least three. For other statistical analysis, a two-sided p value < 0.05 was considered as significant. Statistical analysis was performed using R 2.14.2 (R Foundation, Vienna, Austria) and Stata 11.0 (Statacorp LP, College Station, Texas).

Results

Study selection

The search of the three databases elicited 2,120 records after removal of duplicate records. Ten studies were included in systematic review, including five on agreement between nontriggered and ECG-triggered CT [17-21] and five on prognostic performance [10,22-25]. Subsequently, meta-analysis was performed in three studies [20, 22, 24] with consistent methodology on agreement between nontriggered and ECG-triggered CT (Figure 1).

Table 1 Characteristics of studies on agreement of calcium scoring between nontriggered and triggered CT

Study, year	Patients, *n*	Men, %	Age, year±SD	Setting of study	Type of participants	CT type	Radiation dose	Slice thickness, mm	
								Nontriggered CT	ECG-triggered CT
Budoff 2011 [17]	50	n/a	n/a	COPD cohort	Invited smokers of >10 pack-years	64-MDCT	Low	2.5	2.5
Einstein 2010 [18]	492	44	n/a	Routine clinical population	Consecutively referred adults	16-SPECT/CT 16-PET/CT 64-PET/CT	Low	n/a	n/a
Kim 2008 [19]	128	100	52±7	Lung cancer screening	Consecutively referred elder smokers	40-MDCT	Low	2.5	2.5
Kirsch 2011 [20]	163	78	51±9	Asymptomatic	Consecutively referred elder adults	16- and 64MDCT	Normal	5.0	3.0
Wu 2008 [21]	483	66	62±13	Lung cancer screening	Self-referred elder adults	16MDCT	Low	3.0	3.0

SD = standard deviation; CT = computed tomography; n/a = not available; COPD = chronic obstructive pulmonary disease; ECG = electrocardiographic; MDCT = multi-detector computed tomography; SPECT = single-photon emission computed tomography; PET = positron emission tomography.

Study characteristics

The systematic review included 35,344 participants (range of mean age, 51 to 65 years), comprising 21,558 (61%) men, 13,736 (39%) women and 50 (0.1%) individuals without indicated gender (Tables 1 and 2). Six (60%) studies were prospective, four (40%) retrospective. Four studies (40%) were from North America, three (30%) from Europe, and three (30%) from Asia. All studies were published in English.

Different CT modalities were utilized, ranging from single-slice to 64-row MDCT. Also CT systems as part of single-photon emission computed tomography (SPECT) / CT and positron emission tomography (PET) / CT were utilized [18]. Low-dose acquisition was applied in eight studies (80%), normal dose in one study (10%). In one (10%) study the radiation dose was not reported. Scan data derived from nontriggered CT were reconstructed with different slice thicknesses, ranging from 1.25 to 10 mm (Tables 1 and 2). The most commonly used slice thicknesses were 2.5 / 3 mm and 5 mm. Four studies used a (medium-) smooth kernel [10,20,21,25], the other studies did not indicate the applied kernel. Six studies utilized Agatston scoring [10,17-19,21,25], while four others utilized visual grading of the presence and extent of coronary calcification. No study utilized contrast media. No phantom study was included.

Table 2 Characteristics of studies on prognostic performance of calcium scoring for cardiovascular death or events

Study, year	Patients, *n*	Men, %	Age, year±SD	Setting of study	Type of participants	CT type	Radiation dose	Slice thickness, mm	Cohort name
Itani 2004 [22]	6120	55	61±11	Lung cancer screening	Invited elder adults	Single-slice CT	Low	10	Nagano
Jacobs 2011 [23]	10410	58	62±12	Routine clinical population	Retrospectively included elder adults	4-, 8-, 16-, 40- and 64-MDCT	n/a	3.0-10	PROVIDI
Jacobs 2012 [10]	7557	83	60±6	Lung cancer screening	Invited elder heavy smokers	16-MDCT	Low	3.1	NELSON
Shemesh 2010 [24]	8782	49	65±7	Lung cancer screening	Invited elder heavy smokers	Single- and multi-slice CT	Low	1.25-5	I-ELCAP
Sverzellati 2012 [25]	1159	68	58±6	Lung cancer screening	Invited elder heavy smokers	16-MDCT	Low	5	MILD

SD = standard deviation; CT = computed tomography; n/a = not available; MDCT = multi-detector computed tomography; PROVIDI = the prognostic value of unrequested information in diagnostic imaging; NELSON = the Dutch-Belgian randomized lung cancer screening trial; I-ELCAP = international early lung cancer action program; MILD = multi-centric Italian lung detection.

Study quality

All five studies on agreement between the nontriggered and ECG-triggered CT were of high quality (score ≥ 10 according to the QUADAS tool). Suboptimal scores were present in two QUADAS items: none of the five studies mentioned uninterpretable results (item 13); three studies did not mention whether there were withdrawals (item 14) (Supplemental Table B).

All five studies on prognostic performance were of high quality (a score ≥ 12 according the quality assessment criteria to evaluate prognosis of coronary calcification). Suboptimal scores were present for criterion 4: none of the five studies reported results by ethnicity (Supplemental Table C).

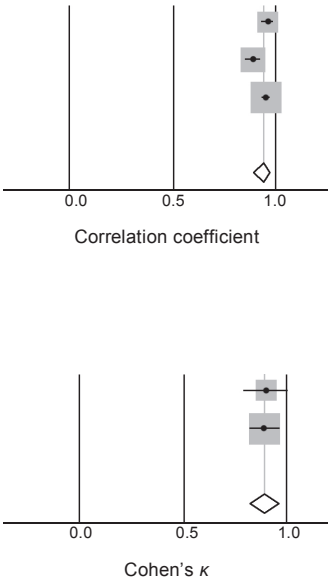
Validation of calcium scoring in nontriggered CT

Five studies were included in the systematic review, comprising 1,316 cardiac asymptomatic participants (Table 1) [17-21]. Diagnostic performance of nontriggered CT was calculated in four studies with 1,153 subjects (Table 3) [17-19,21], in which 137 (11.9%) had CS of 100 - 400 in nontriggered CT. Fifty-five subjects (8.8%) showed no coronary calcification in nontriggered CT examination among 625 subjects with CS > 0 in ECG-triggered CT. In those fifty-five subjects, fifty-two (8.3%) had CS 1-100 in ECG-triggered CT, and three (0.5%) had CS 100-400. Among 162 subjects with CS ≥ 400 in ECG-triggered CT, nontriggered CT underestimated the CS in 31 subjects (19.1%). In these thirty-one subjects, two (1.2%) had CS 1 - 100 in nontriggered CT, and twenty-nine (17.9%) had CS 100-400. On the other hand, among

991 subjects with CS < 400 in ECG-triggered CT, nontriggered CT showed a CS ≥ 400 in twenty-six subjects (2.6%) and thus overestimated the CS. In those twenty-six subjects, one (0.1%) had CS 1-100 in ECG-triggered CT, and twenty-five (2.5%) had CS 100 - 400.

Figure 2 Forest plots for agreement of coronary calcium scoring between nontriggered and electrocardiography-triggered computed tomography

Study, year	Effect size (95% CI)	Size,*n*	Forest plots
Calcium score			
	Correlation coefficient		
Budoff 2011[17]	0.96 (0.93, 0.98)	50	
Kim 2008[19]	0.89 (0.85, 0.92)	128	
Wu 2008[21]	0.95 (0.94, 0.96)	483	
Pooled	0.94 (0.89, 0.97)	661	
Heterogeneity	p(Q) < 0.001, I^2 = 89.1%		
Overall effect	p(Z) < 0.001		
Publication bias	p(B) = 0.60, p(E) = 0.80		
4 categories of calcium score*			
	Concordance Cohen's κ		
Budoff 2011[17]	94% 0.90 (0.79, 1.00)	50	
Wu 2008[21]	93% 0.89 (0.82, 0.96)	483	
Pooled	0.89 (0.83, 0.95)	533	
Heterogeneity	p(Q) = 0.88, I^2 = 0%		
Overall effect	p(Z) < 0.001		



*The four categories of calcium score were defined as 0, 1-99, 100-399, and ≥ 400. p(Q) = p value for Q statistic; p(Z) = p value for Z test; p(B) = p value for Begg and Mazumdar rank correlation test; p(E) = p value for Egger's regression test.

Meta-analysis was performed in three studies comprising 661 participants (Figure 2) [17,19,21]. The study by Kirsch could not be included because it evaluated the amount of coronary calcification using visual grading score [20]. The pooled CC for CS was 0.94 (95%CI: 0.89 to 0. 97). The pooled Cohen's κ was 0.89 (95%CI: 0.83 to 0.95) for four categories of the calcium score. Heterogeneity was found in the pooling calculation of the calcium score (P for Q statistic < 0.001 and I^2 >50%). No publication bias was found in the pooling calculation of the calcium score (P > 0.05). Publication bias testing was not performed in the pooling calculation of four CS categories, due to insufficient number of studies.

Prognosis of calcium scoring in nontriggered CT

Five studies were included, comprising 34,028 cardiac asymptomatic participants (Table 2) [10,22-25]. In the five studies, mean follow-up duration was 45 months (range, 10 to 72 months). None of the participants in the included studies had a history or symptoms of cardiovascular diseases before CT examination. During follow-up, 207 cardiovascular deaths and 675 cardiovascular events were observed. Overall, with increasing CS categories, increasing unadjusted and adjusted hazard ratio for cardiovascular death or events was observed. Risks in CS categories were not consistently reported, however in one study, unadjusted and adjusted hazard ratio increased up to 7.5 and 5.3 for CS > 1000, respectively (Table 4) [10].

A small percentage of subjects with zero CS in nontriggered CT suffered cardiovascular death or events. During a mean follow-up of 45 months, 47 cardiovascular deaths (0.55%) were found in 8,487 subjects with zero CS [22,24], whereas 72 cardiovascular events (1.3%) occurred in 5,249 subjects with zero CS [10,23]. However, the event rate for subjects with positive CS was higher. During follow-up, 160 cardiovascular deaths (2.5%) were found in 6,415 subjects with positive CS [22,24], whereas 570 cardiovascular events (4.5%) occurred in 12,718 subjects with positive CS [10,23].

Discussion

In this systematic review and meta-analysis, we aimed to investigate whether coronary calcium scoring can be performed in nontriggered thoracic CT, for instance utilized in lung cancer screening. A strong correlation in calcium score categories between nontriggered and ECG-triggered CT was found. In cardiac asymptomatic elderly and smokers, mainly from lung cancer screening trials, increasing coronary calcium burden translated into a higher risk of cardiovascular death or events.

Calcium score for individual atherosclerotic lesions is greatly influenced by motion [12]. Regardless, we found that the correlation in CS between nontriggered and ECG-triggered CT was excellent (r = 0.94) on a group level. In broad CS categories, we found a high agreement between nontriggered and ECG-triggered CT (Cohen's κ = 0.89). Thus, for an individual patient, although variability in CS between nontriggered and ECG-triggered CT is likely considerable, broad CS categories can potentially be utilized for cardiovascular risk stratification [3,4].

Table 3 Agreement of coronary calcium scoring between nontriggered thoracic and ECG-triggered cardiac CT, and diagnostic performance of calcium scoring on nontriggered CT

	Agreement			Diagnostic performance*		
Study, year	Scoring in nontriggered CT	Reference scoring in triggered CT	Agreement between nontriggered and triggered CT	False negative calcium score, %	Underestimated high calcium score, %	Overestimated high calcium score, %
Budoff 2011 [17]	CS	CS	r = 0.96	0%	0%	8.6%
	4 categories of CS†	4 categories of CS†	κ = 0.90, concordance=94%			
Einstein 2010 [18]	6 categories of CS‡	6 categories of CS‡	κ = 0.89, concordance=63%	14.0%	23.4%	4.9%
Kim 2008 [19]	CS	CS	r = 0.89	9.3%	0%	0%
Kirsch 2011 [20]	Visual grading score§	CS	r = 0.83	n/c	n/c	n/c
Wu 2008 [21]	CS	CS	r = 0.95	2.3%	15.2%	0.9%
	4 categories of CS†	4 categories of CS†	κ = 0.89, concordance=93%			

*False negative calcium score is indicated as the percentage of CS = 0 subjects on nontriggered CT among CS > 0 subjects on triggered CT. Underestimated high-risk calcium score is indicated as the percentage of CS < 400 subjects on nontriggered CT among CS ≥ 400 subjects on triggered CT.

Overestimated high-risk calcium score is indicated as the percentage of CS ≥ 400 subjects on non-triggered CT among CS < 400 subjects on triggered CT.

†Four categories of CS were defined as 0, 1-99, 100-399, and ≥400.

‡Six categories of CS were defined as 0, 1-9, 10-99, 100-399, 400-999 and ≥1000.

§A score was assigned for each major coronary artery. 0: No calcification; 1: single pixel calcification; 3: dense calcification with blooming artifact; 2: calcification between 1 and 3. The visual grading score (range 0 to 12) was calculated by the sum of the score for each artery. CT = computed tomography; CS = calcium score; n/c = not calculated because different scoring systems were used in nontriggered and ECG-triggered CT.

Absence of coronary calcification in ECG-triggered CT is associated with a very low cardiovascular risk, and thus is commonly utilized to rule out coronary artery disease [3,26]. We found that nontriggered CT can yield a false-negative CS in about 9% of individuals compared to ECG-triggered CT. Furthermore, we found that a zero CS in nontriggered CT indicates a low cardiovascular risk, although nontriggered CT cannot reliably exclude coronary calcification. When a high CS (≥ 400) is found in asymptomatic individuals, the risk of cardiovascular events is elevated. The ACCF / AHA consensus document suggests to regard these individuals as candidates for intensive preventive therapies [5]. The probability of overestimating the CS is low, and thus, it is reasonable to assume an elevated cardiovascular risk in case of a CS ≥ 400 in nontriggered CT. On the other hand, nontriggered CT underestimated the CS in a nonnegligible percentage of individuals with CS ≥ 400 in ECG-triggered CT, thus underestimating cardiovascular risk. In the validation study, 11.9% had a CS of 100-400 in the nontriggered CT examination. In this relatively small proportion of the included study populations, dedicated ECG-triggered CT could be considered, to assess whether the CS is actually ≥ 400. This proportion is much lower than the population percentage in which calcium scoring could be considered according to current consensus documents (40%) [27].

In this study, hazard ratios of CS categories for cardiovascular events were generally lower than in a previous systematic review on calcium scoring derived from ECG-triggered CT [5]. For example, adjusted HR for cardiovascular events was up to 5.3 for CS > 1000 in our study, lower than 10.8 in a previous report in ECG-triggered CT in an elderly population [28]. The relative risk is usually based on the risk of subjects without coronary calcium at baseline as reference category. During a mean follow-up of 45 months, we found that 1.3% subjects without coronary calcium had a cardiovascular event. In contrast, a meta-analysis by Sarwar et al. on ECG-triggered CT reported only 0.47% of subjects without coronary calcium suffered a cardiovascular event during a mean follow-up of 50 months [26]. In that meta-analysis, studies mainly consisted of middle-aged individuals at low-to-intermediate cardiovascular risk, referred for cardiovascular risk evaluation. In contrast,

the majority of the populations in the prognostic studies on calcium scoring using nontriggered CT comprised participants of lung cancer screening trials. The generally higher age and heavier smoking history in the prognostic studies included in our study likely at least partly explain the higher event rate in individuals with zero CS. Besides, this higher event rate in case of zero CS for nontriggered CT may also be explained by the fact that a proportion of the individuals without coronary calcification on the nontriggered CT, actually have a positive CS in ECG-triggered CT. As this reference risk is higher, the relative risk for increasing CS categories also yield lower values. Our finding does suggest that presence of coronary calcification in nontriggered CT is an independent predictor of cardiovascular events. Also, we found that higher calcium burden translated into a higher cardiovascular risk in a large aggregated sample.

Reproducibility of calcium scoring in repeated nontriggered CT has been investigated. Jacobs et al investigated 584 subjects who underwent two nontriggered examinations of the thorax, and calculated the CS for both exams [29]. The calcium scores were divided into the commonly used categories of 0, 1 - 100, 101 - 400, > 400. In 440 cases (75%), the calcium scores of the two CT examinations fell in the same category. In 138 subjects (24%), calcium scores differed by one category, and in 6 subjects (1%) by more than one category. The intra-class correlation coefficient was 0.94. On the other hand, reproducibility of calcium scoring in ECG-triggered CT is also not perfect. Using ECG-triggered CT, Rutten reported that 76% to 85% of individuals ended up in the same CS category, and in 15% to 24%, the results differed by one category [30].

The agreement of repeated calcium scoring in nontriggered CT within and between observers is high, although slightly lower than in ECG-triggered CT. Nearly all studies in this systematic review investigated either intra- or inter-observer variability of calcium scoring in nontriggered CT. For example, in 483 subjects, Wu reported an inter-observer variability of 3.6% for ECG-triggered CT, and of 9.6% for nontriggered CT [21]. However, all studies found a very strong concordance in score categorization within and between observers (kappa values 0.77 to 0.91, intraclass correlation coefficient 0.93 to 0.99) [10,18,20,21,23-25].

The majority of included studies (80%) were based on low-dose thoracic CT [10,17-19,21,22,24,25], which has a lower radiation dose than a dedicated cardiac CT for calcium scoring. A typical effective radiation dose for low-dose CT utilized in lung cancer screening is 0.8 - 0.9 mSv for normal sized body [8,21]. However, the mean dose for a cardiac CT for calcium scoring is approximately 1.0 - 2.9 mSv, depending on scanner type and scanning protocol [21,31].

Table 4 Prognostic performance of coronary calcium scoring for cardiovascular death or events in nontriggered CT

Study, year	Follow-up, months (range)	Endpoint event, *n*	Calcium scoring method in non-triggered CT	Calcium scoring cut-off	Event number/category number, percentage	Unadjusted hazard ratio of calcium score (95%CI)	Adjusted hazard ratio of calcium score(95%CI)	Adjusted for factors
Itani 2004 [22]	48	Cardiac death, 14	Presence of coronary calcification	Zero calcium score	4/4914,0.08%	1.0(reference)	n/a	n/a
				Positive calcium score	10/1206,0.83%	2.7(0.8, 9.4)		
Jacobs 2011 [23]	18	Cardiovascular event, 515	4 categories of visual grading score*	Zero calcium score	62/3435,1.8%	1.0(reference)	1.0(reference)	Age, sex, indication for CT, image quality and type of medical center
				Visual score:1-2	113/2498,4.5%	2.8(2.0, 3.8)	2.2(1.6, 3.0)	
				Visual score:3-5	149/2603,5.7%	3.8(2.9, 5.2)	2.5(1.8, 3.4)	
				Visual score:6-12	191/1874,10.2%	6.9(5.2, 9.2)	3.7(2.7, 5.2)	
Jacobs 2012 [10]	10(1-21)	Cardiovascular event, 127	4 categories of calcium score	Zero calcium score	10/1814,0.6%	1.0(reference)	1.0(reference)	Age, sex, smoking status, hypertension, hypercholesterolemia and diabetes
				Calcium score:1-100	27/2191,1.2%	1.9(0.9, 4.2)	1.8(0.8, 3.9)	
				Calcium score:101-1000	32/2267,1.4%	2.2(1.0, 4.8)	1.9(0.9, 4.2)	
				Calcium score:>1000	58/1285,4.5%	7.5(3.6, 15.7)	5.3(2.5, 11.6)	
Shemesh 2010 [24]	72(1-92)	Cardiovascular death, 193	3 categories of visual grading score†	Zero calcium score	43/3573,1.2%	1.0(reference)	1.0(reference)	Age, sex and smoking pack-years
				Visual score:1-3	66/3569,1.8%	1.6(1.1, 2.4)	1.0(0.7, 1.5)	
				Visual score:4-12	84/1640,5.1%	4.7(3.6, 6.8)	2.1(1.4, 3.1)	
Sverzellati 2012 [25]	36(1-54)	Cardiovascular event, 33	2 categories of calcium score	Calcium score:≤400	26/1079,2.4%	n/a	1.0(reference)	Age, sex, diabetes, hypertension, smoking status and smoking duration
				Calcium score:>400	7/80,8.8%		2.9(1.1, 7.3)	

*A grading score was assigned for each major coronary artery. 0: No calcification. 1: One or two calcifications. 2: >2 calcifications or one calcification extending ≥2 slices. 3: Calcification covering a large coronary segment. Four visual grades were stratified by the sum of the score (0, 1-2, 3-5 and 6-12).

†A grading score was assigned for each major coronary artery. 0: No calcification. 1: ≤⅓ of the artery length showed calcification. 2: ⅓ to ⅔. 3:≥⅔. Three visual grades were stratified by the sum of the score (0, 1-3 and 4-12). CI = confidence interval; n/a = not available.

Clinical implication

A large number of nontriggered CT examinations are annually performed world-wide. In the aging and smoking population, coronary calcification is a common finding. A lung cancer screening trial reported that over 70% of the participants had coronary calcification [10]. The group at risk for lung cancer overlaps with the group at highest risk of cardiovascular diseases, because at least aging and smoking are two major risk factors for both diseases. There may be a substantial primary prevention potential if the calcium score can be derived from the same examination, at least in participants of lung cancer screening trials. While results from the one study in a clinical population suggest that the extent of coronary calcification is also predictive outside lung cancer screening setting, more studies are needed to confirm the value of calcium scoring in routine clinical thoracic CT

We observed that CS categorization between nontriggered and ECG-triggered CT correlated very well, and increasing CS categories based on nontriggered CT are predictive of increasing cardiovascular risk. Thus, for subjects who were examined by nontriggered thoracic CT, the cardiovascular risk could potentially be stratified by performing calcium scoring. Subjects identified in nontriggered CT as having high CS could be considered as candidates of intensive risk factor modification, especially in an aging and smoking population such as the participants in lung cancer screening. However, a zero calcium score in nontriggered CT does not exclude coronary calcification.

Furthermore, cardiovascular event rate of subjects without CS in nontriggered CT is higher than in ECG-triggered CT. Absent coronary calcification in nontriggered CT may not reliably exclude the risk of cardiovascular events. Future studies on this topic are needed to provide stronger support for coronary calcium scoring in nontriggered CT.

Limitations

Firstly, despite our favorable results it remains to be clarified whether differences in the accuracy between nontriggered and ECG-triggered CS measures translate into differences in prognostic value. For example, a zero CS in nontriggered CT may render a positive CS in ECG-triggered exams. Secondly, for agreement in calcium scoring between nontriggered and ECG-triggered CT, the number of studies and participants in the meta-analysis was

relatively low. To compare calcium scoring between nontriggered and ECG-triggered CT, the patients have to be scanned twice in a short time-interval, and at doubled radiation dose. This could contribute to the relatively small number of studies in this field. Our conclusions are not only based on the pooling calculations but also on the systematic review. Besides, for the second part of our study, regarding the prognostic value, the aggregated sample size was over 30,000 individuals. Thirdly, different calcium scoring methods were utilized in studies on prognostic performance. Meta-analysis could therefore not be performed to assess predictive value of the calcium score derived from nontriggered CT. However, heavier calcium burden was in the systematic review associated with increasing cardiovascular risk. Finally, included studies were fairly heterogeneous in terms of participant population, imaging equipment and acquisition protocol. Those factors weakened the strength of meta-analysis. We used a random effects model to compensate for at least some of the heterogeneity in the pooling calculation. On the other hand, the differences in imaging procedures also reflect the heterogeneity of procedures in clinical practice. Despite different CT equipment and calcium scoring methods, at least the presence or absence of coronary calcium is clear. Results on the presence and absence of coronary calcification should not differ significantly based on important CS categories. Thus, our conclusions based on findings related to those points are solid.

Conclusions

In this systematic review and meta-analysis, strong agreement in CS categorization was found between nontriggered CT and ECG-triggered CT. Compared to ECG-triggered CT, a high calcium score category in nontriggered CT is a fairly reliable finding. However, nontriggered CT yielded false-negative CS in 8.8% of individuals, and underestimated high CS in 19.1%. In cardiac asymptomatic participants mainly from lung cancer screening trials, increasing CS categories in nontriggered CT were associated with increasing risk of cardiovascular events. Our analysis presents preliminary evidence for the prognostic value and potential role of calcium scoring in nontriggered CT. However, it does not suggest that nontriggered examinations can replace dedicated, ECG- triggered CT.

References

1. Agatston AS, Janowitz WR, Hildner FJ, et al. Quantification of coronary artery calcium using ultrafast computed tomography. J Am Coll Cardiol. 1990;15:827-832.
2. Nasir K, Rubin J, Blaha MJ, et al. Interplay of coronary artery calcification and traditional risk factors for the prediction of all-cause mortality in asymptomatic individuals. Circulation-Cardiovascular Imaging. 2012;5:467-473.
3. Oudkerk M, Stillman AE, Halliburton SS, et al. Coronary artery calcium screening: current status and recommendations from the European Society of Cardiac Radiology and North American Society for Cardiovascular Imaging. European radiology. 2008;18:2785-2807.
4. Elias-Smale SE, Proenca RV, Koller MT, et al. Coronary calcium score improves classification of coronary heart disease risk in the elderly: the Rotterdam study. J Am Coll Cardiol. 2010;56:1407-1414.
5. Greenland P, Bonow RO, Brundage BH, et al. ACCF/AHA 2007 clinical expert consensus document on coronary artery calcium scoring by computed tomography in global cardiovascular risk assessment and in evaluation of patients with chest pain: a report of the American College of Cardiology Foundation Clinical Expert Consensus Task Force (ACCF/AHA Writing Committee to Update the 2000 Expert Consensus Document on Electron Beam Computed Tomography). Circulation. 2007;115:402-426.
6. Jain A, McClelland RL, Polak JF, et al. Cardiovascular Imaging for Assessing Cardiovascular Risk in Asymptomatic Men Versus Women The Multi-Ethnic Study of Atherosclerosis (MESA). Circulation-Cardiovascular Imaging. 2011;4:8-15.
7. Berrington de Gonzalez A, Mahesh M, Kim KP, et al. Projected cancer risks from computed tomographic scans performed in the United States in 2007. Arch Intern Med. 2009;169:2071-2077.
8. van Klaveren RJ, Oudkerk M, Prokop M, et al. Management of lung nodules detected by volume CT scanning. The New England journal of medicine. 2009;361:2221-2229.
9. Oei HH, Vliegenthart R, Hofman A, Oudkerk M, Witteman JC. Risk factors for coronary calcification in older subjects. The Rotterdam Coronary Calcification Study. European heart journal. 2004;25:48-55.
10. Jacobs PC, Gondrie MJA, van der Graaf Y, et al. Coronary artery calcium can predict all-cause mortality and cardiovascular events on low-dose CT screening for lung cancer. AJR American journal of roentgenology. 2012;198:505-511.
11. Mets OM, de Jong PA, Prokop M. Computed tomographic screening for lung cancer: an opportunity to evaluate other diseases. JAMA : the journal of the American Medical Association. 2012;308:1433-1434.

12. Greuter MJW, Groen JM, Nicolai LJ, Dijkstra H, Oudkerk M. A model for quantitative correction of coronary calcium scores on multidetector, dual source, and electron beam computed tomography for influences of linear motion, calcification density, and temporal resolution: A cardiac phantom study. Medical Physics. 2009;36:5079-5088.
13. White CS, Jerome S. Coronary calcium scoring on nongated chest CT: is it ready for prime time? Journal of cardiovascular computed tomography. 2011;5:119-121.
14. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. British Medical Journal. 2009;339:332-336.
15. Prokop M. New challenges in MDCT. European radiology. 2005;15 Suppl 5:E35-45.
16. Whiting PF, Weswood ME, Rutjes AW, et al. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. BMC Medical Research Methodology. 2006;6:9.
17. Budoff MJ, Nasir K, Kinney GL, et al. Coronary artery and thoracic calcium on noncontrast thoracic CT scans: comparison of ungated and gated examinations in patients from the COPD Gene cohort. Journal of cardiovascular computed tomography. 2011;5:113-118.
18. Einstein AJ, Johnson LL, Bokhari S, et al. Agreement of visual estimation of coronary artery calcium from low-dose CT attenuation correction scans in hybrid PET/CT and SPECT/CT with standard Agatston score. Journal of the American College of Cardiology. 2010;56:1914-1921.
19. Kim SM, Chung MJ, Lee KS, et al. Coronary calcium screening using low-dose lung cancer screening: effectiveness of MDCT with retrospective reconstruction. AJR. American journal of roentgenology. 2008;190:917-922.
20. Kirsch J, Buitrago I, Mohammed TL, et al. Detection of coronary calcium during standard chest computed tomography correlates with multi-detector computed tomography coronary artery calcium score. The international journal of cardiovascular imaging. 2012;28:1249-1256.
21. Wu MT, Yang P, Huang YL, et al. Coronary arterial calcification on low-dose ungated MDCT for lung cancer screening: concordance study with dedicated cardiac CT. AJR. American journal of roentgenology. 2008;190:923-928.
22. Itani Y, Sone S, Nakayama T, et al. Coronary artery calcification detected by a mobile helical computed tomography unit and future cardiovascular death: 4-year follow-up of 6120 asymptomatic Japanese. Heart and vessels. 2004;19:161-163.
23. Jacobs PC, Gondrie MJ, Mali WP, et al. Unrequested information from routine diagnostic chest CT predicts future cardiovascular events. European radiology. 2011;21:1577-1585.

24. Shemesh J, Henschke CI, Shaham D, et al. Ordinal scoring of coronary artery calcifications on low-dose CT scans of the chest is predictive of death from cardiovascular disease. Radiology. 2010;257:541-548.
25. Sverzellati N, Cademartiri F, Bravi F, et al. Relationship and prognostic value of modified coronary artery calcium score, FEV1, and emphysema in lung cancer screening population: the MILD trial. Radiology. 2012;262:460-467.
26. Sarwar A, Shaw LJ, Shapiro MD, et al. Diagnostic and prognostic value of absence of coronary artery calcification. JACC. Cardiovascular imaging. 2009;2:675-688.
27. Greenland P, Gaziano JM. Selecting asymptomatic patients for coronary computed tomography or electrocardiographic exercise testing. New Engl J Med. 2003;349:465-473.
28. Vliegenthart R, Oudkerk M, Hofman A, et al. Coronary calcification improves cardiovascular risk prediction in the elderly. Circulation. 2005;112:572-577.
29. Jacobs PC, Isgum I, Gondrie MJ, et al. Coronary artery calcification scoring in low-dose ungated CT screening for lung cancer: interscan agreement. AJR. American journal of roentgenology. 2010;194:1244-1249.
30. Rutten A, Isgum I, Prokop M. Calcium scoring with prospectively ECG-triggered CT: using overlapping datasets generated with MPR decreases inter-scan variability. European journal of radiology. 2011;80:83-88.
31. Dey D, Nakazato R, Pimentel R, et al. Low radiation coronary calcium scoring by dual-source CT with tube current optimization based on patient body size. Journal of cardiovascular computed tomography. 2012;6:113-120.

Supplemental Table A Literature search strategy

Search terms used to identify relevant citations

PubMed
("Tomography, X-Ray Computed"[MeSH] OR "computed tomography"[tiab] OR CT[tiab] OR "MDCT") AND ("untriggered" OR "ungated" OR "nongated" OR "nontriggered" OR "nonelectrocardiogram" OR "thorax"[MeSH] OR "chest" OR "thoracic" OR "lung" OR "pulmonary" OR "torso") AND ("coronary vessels"[MeSH] OR "Coronary") AND ("Calcium" OR "calcification" OR "calcific" OR "calcified") AND 1900/01:2012/11[dp]

EmBase
#1: ((Computed tomography) OR CT:ab,ti OR MDCT) AND (untriggered OR ungated OR nongated OR nontriggered OR nonelectrocardiogram OR thorax OR chest OR thoracic OR lung:ab,ti OR pulmonary:ab,ti OR torso) AND Coronary AND (Calcium OR calcification OR calcific OR calcified)
Grammar in advanced search: #1 AND [1-1-1900]/sd NOT (#1 AND [30-11-2012]/sd)

Web of Knowledge
#1 topic: ((Computed tomography) OR CT OR MDCT)
#2 topic: (untriggered OR ungated OR nongated OR nontriggered OR nonelectrocardiogram OR thorax OR chest OR thoracic OR lung OR pulmonary OR torso)
#3 topic: Coronary
#4 topic: (Calcium OR calcification OR calcific OR calcified)
Grammar: #1 topic and #2 topic and #3 topic and #4 topic

Supplemental Table B Quality assessment for validation studies on agreement and diagnostic performance, by the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool

	Item1: Representative patient sample	Item2: Selection criteria clearly described	Item3: Acceptable reference standard	Item4: Acceptable delay between tests*	Item5: Partial verification avoided	Item6: Differential verification avoided	Item7: Incorporation avoided	Item8: Adequate index test description	Item9:Adequate reference standard description	Item10: Index test blinded to reference standard	Item11: Reference standard blinded to index test	Item12: Clinical data available as in practice	Item13: Uninterpretable test results reported	Item14: Withdrawals explained	Score†
Budoff 2011[17]	+	+	+	+	+	+	+	+	+			+			12.0
Einstein 2010[18]	+	+	+	+	+	–	+	+	+	+	+	+			12.0
Kim 2008[19]	+	+	+	+	+	+	+	+	+			+		+	12.5
Kirsch 2011[20]	+	+	+	+	+	+	+	+	+	+	+	+			13.0
Wu 2008[21]	+	+	+	+	+	+	+	+	+	+	+	+		+	13.5

*Maximum delay of 2 months between nontriggered and reference examination was considered as acceptable.

†For each study, a quality score was accumulated by assigning 1 point to "yes" item, 0.5 point to "unclear" item, and 0 to "no" item. The total possible score was 14. A score of ≥ 10 points was considered as high quality, and a score between 6 and 9 points as moderate quality, a score of ≤ 5 as low quality. "+" = yes; "-" = no; empty = unclear.

Supplemental Table C Quality assessment for studies on prognosis, by the quality assessment criteria of prognostic studies on coronary artery calcium in American College of Cardiology Foundation / American Heart Association (ACCF/AHA) guideline

	Criterion 1: Retrospective vs. prospective study	Criterion 2: Potential for referral bias	Criterion 3: Reporting coronary calcification by CHD death or myocardial infarction	Criterion 4: Reporting of results by gender or ethnicity	Criterion 5: Sample size greater than 1000	Criterion 6: Potential for limited challenge	Criterion 7: Risk factor reporting	Criterion 8: Covariate or risk-adjusted outcomes	Score*
Itani 2004[22]	2	1	2	1	1	1	3	0	11
Jacobs 2011[23]	1	2	2	0	1	2	2	1	11
Jacobs 2012[10]	2	2	2	0	1	2	3	1	13
Shemesh 2010[24]	2	2	2	1	1	2	3	1	14
Sverzellati 2012[25]	2	2	2	1	1	2	3	1	14

*For each study, a quality score was accumulated by assigning a score for each criterion as the following: $_{Criterion}$1: Retrospective vs. prospective study (1 = retrospective, 2 = prospective). $_{Criterion}$2: Potential for referral bias (0 = clinically referred patients, 1 = unselected cohort, 2 = population sample). $_{Criterion}$3: Reporting coronary calcification by CHD death or myocardial infarction (1 = no, 2 = yes). $_{Criterion}$4: Reporting of results by gender or ethnicity (0 = no, 1 = gender only, 2 = ethnicity only, 3 = both). $_{Criterion}$5: Sample size ≥ 1000 (0 = no, 1 = yes). $_{Criterion}$6: Potential for limited challenge (1 = no reporting of calcium outcomes in low- to high-risk global risk scores, 2 = reporting of calcium outcomes in low- to high-risk global risk scores). $_{Criterion}$7: Risk factor reporting (1 = historical only, 2 = measured in subset, 3 = measured in all subjects). $_{Criterion}$8: Covariate or risk-adjusted outcomes (0 = no, 1 = yes).

Total possible score was 16. A score of ≥ 11 points was considered as high quality, and a score between 7 and 10 points as moderate quality, a score of ≤ 6 as low quality. CHD = coronary heart disease.

Chapter 7

Can Nontriggered Thoracic CT be Used for Coronary Artery Calcium Scoring? A Phantom Study

Medical Physics. 2013 Jul 18 [Epub ahead of print]

Xueqian Xie[12] / Marcel J.W. Greuter[1] / Jaap M. Groen[3] / Geertruida H. de Bock[4]
Matthijs Oudkerk[2] / Pim A. de Jong[5] / Rozemarijn Vliegenthart[12]

University of Groningen, University Medical Center Groningen, [1]Department of Radiology,
[2]Center for Medical Imaging – North East Netherlands, [4]Department of Epidemiology;
[3]Department of Radiology, Zaans Medical Center;
[5]Department of Radiology, University Medical Center Utrecht

Abstract

Background

Coronary artery calcium score, traditionally based on electrocardiography (ECG)-triggered computed tomography (CT), predicts cardiovascular risk. However, nontriggered CT is extensively utilized. The study-purpose is to evaluate the in-vitro agreement in coronary calcium score between nontriggered thoracic CT and ECG-triggered cardiac CT.

Materials and Methods

Three artificial coronary arteries containing calcifications of different densities (high, medium and low), and sizes (large, medium and small), were studied in a moving cardiac phantom. Two 64-detector CT systems were used. The phantom moved at 0 to 90 mm/s in nontriggered low-dose CT as index test, and at 0 to 30 mm/s in ECG-triggered CT as reference. Differences in calcium scores between nontriggered and ECG-triggered CT were analyzed by *t*-test and 95% confidence interval. The sensitivity to detect calcification was calculated as the percentage of positive calcium scores.

Results

Overall, calcium scores in nontriggered CT were not significantly different to those in ECG-triggered CT ($p > 0.05$). Calcium scores in nontriggered CT were within the 95% confidence interval of calcium scores in ECG-triggered CT, except predominantly at higher velocities (≥ 50 mm/s) for the high-density and large-size calcifications. The sensitivity for a nonzero calcium score was 100% for large calcifications, but 46 ± 11% for small calcifications in nontriggered CT.

Conclusions

When performing multiple measurements, good agreement in positive calcium scores is found between nontriggered thoracic and ECG-triggered cardiac CT. Agreement decreases with increasing coronary velocity. From this phantom study, it can be concluded that a high calcium score can be detected by nontriggered CT, and thus, that nontriggered CT likely can identify individuals at high risk of cardiovascular disease. On the other hand, a zero calcium score in nontriggered CT does not reliably exclude coronary calcification.

Introduction

Coronary artery calcium score, derived from computed tomography (CT) and traditionally expressed as Agatston score, reflects the extent of coronary atherosclerosis. Calcium score is an independent predictor of cardiovascular morbidity and mortality [1,2]. Recent guidelines from cardiological and radiological scientific societies support the utilization of calcium score to assess cardiovascular risk in asymptomatic individuals at intermediate risk [3,4]. CT is a well-accepted imaging modality to assess calcium score [4]. Due to periodic and rapid movement of coronary arteries, prospective or retrospective electrocardiographic (ECG)-triggering or gating has been employed in CT to allow least-motion imaging of coronary arteries with calcified lesions. Because of the relatively lower radiation dose, prospective ECG-triggering is the current standard practice [5].

So far, ECG-triggering has been part of the standard CT protocol for coronary calcium scoring. There may be substantial unused potential if calcium score could also be determined in nontriggered thoracic CT. The latter procedure is extensively utilized worldwide. For example, 13.6 million nontriggered thoracic CT examinations were performed in the United States in 2007, in contrast to 0.7 million ECG-triggered CT examinations for calcium scoring [6]. Interests in coronary calcium quantification based on thoracic CT examinations have further increased in view of the recent evidence on reduced mortality with lung cancer CT screening [7]. If the same screening examination could be performed to assess early-stage lung cancer as well as coronary calcium score, the benefits and cost-efficiency of low-dose thoracic CT might considerably increase. However, variability in coronary calcium quantification between those two techniques was substantial. To determine the influence of variability on the correlation between those two techniques, an in-depth validation study for the impact of motion artifacts on calcium score in nontriggered CT using a phantom model has to be performed.

The purpose of this study was therefore to evaluate the agreement in coronary calcium scoring between low-dose thoracic computed tomography (CT) and dedicated cardiac CT examinations with ECG-triggering using a moving cardiac phantom.

Materials and Methods

Phantom

The moving cardiac phantom (Sim2D, QRM, Möhrendorf, Germany) consisted of a computer-controlled motion unit and a water container. The cardiac phantom was placed inside a thoracic phantom (Thorax, QRM, Möhrendorf, Germany) to simulate thoracic environment (Figure 1). A lever on the motion unit moved at a linear velocity programmed at 0

to 90 mm/s (Figure 2), which is within the velocity range of in-vivo coronary arteries [8]. ECG signal was simulated, and synchronized with the lever movement cycle.

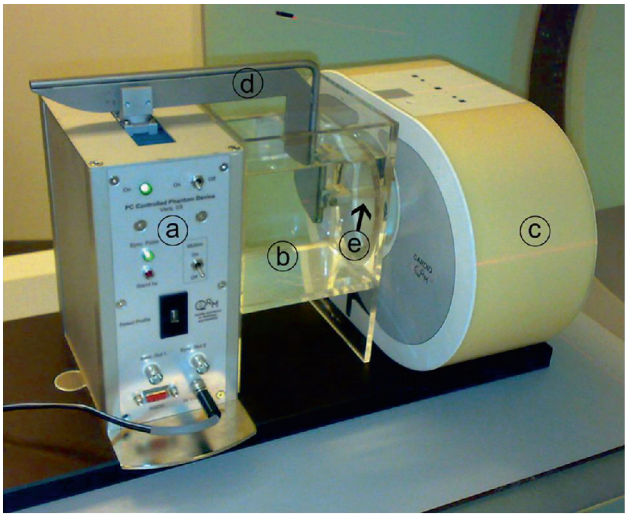


Figure 1 The moving cardiac phantom, comprising of a computer controlled motion unit (**a**), a water container (**b**), a thoracic phantom (**c**), a lever (**d**), and an artificial coronary artery (**e**).

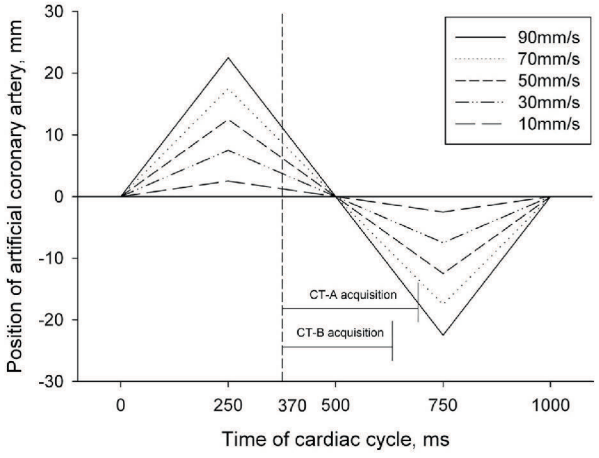


Figure 2 Position of the artificial coronary artery as a function of time, for velocities of 10, 30, 50, 70 and 90 mm/s. The image acquisition windows of CT-A and CT-B in ECG-triggered acquisition mode are indicated, starting at 370 ms and with a width of 330 ms and 267 ms respectively.

Three artificial coronary arteries with calcified lesions made of hydroxyapatite were investigated (Figure 3). The density of the calcified lesions differed per artificial artery (high density, 800 Hounsfield units (HU), medium density, 400 HU and low density, 200 HU, all at 120 kV). Each artificial artery contained three calcifications of different size (large, 62.8 mm^3, medium, 24.6 mm^3, and small, 9.1 mm^3) (Table 1). The density and size of the calcified lesions resembled the density and size of in-vivo coronary calcifications

[9,10]. The background density of the artificial arteries was 50 HU at 120 kV, made of polyurethane resin, resembling in-vivo CT attenuation of blood. The artificial arteries were attached to the lever and imaged one by one to avoid overlap artifacts. The lever with one attached artificial artery was placed inside the water container which was subsequently inserted in the thoracic phantom. The artificial artery was positioned parallel to the central axis of the CT examination bed (*z*-axis) and moved in the *x*-*y* plane (perpendicular to the *z*-axis).

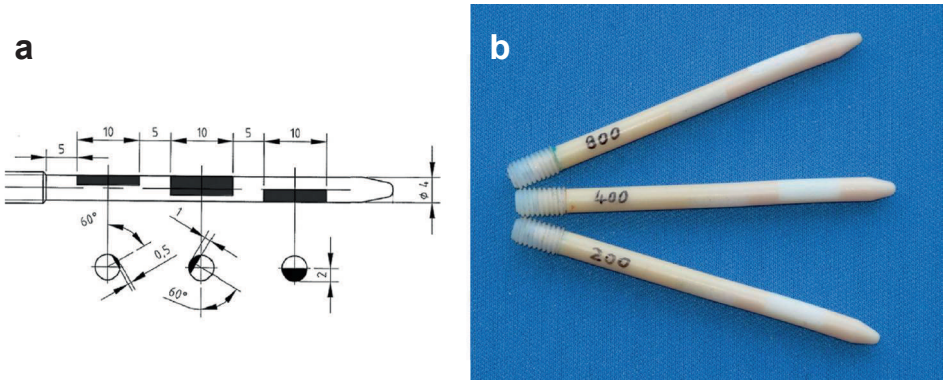


Figure 3 Artificial arteries with calcified lesions of high (800 HU), medium (400 HU) and low density (200 HU) in large, medium and small size. Schematic drawing (**a**) of an artificial artery, the black shadows indicated the artificial calcified lesions in the artificial artery, the length, radius and rotation angles of the artificial calcified lesions are also indicated. A photograph (**b**) of three artificial arteries.

CT imaging

Nontriggered thoracic and ECG-triggered cardiac CT examinations were performed in two medical centers using 64-slice CT systems from two different vendors (CT-A: Soma-tom Sensation 64, Siemens, Forchheim, Germany; CT-B: Brilliance 64, Philips, Best, The Netherlands). The acquisition protocol for the nontriggered examination was derived from the Dutch-Belgian randomized lung cancer screening trial (NELSON) [11]. The ECG-triggered CT examination was performed according to the daily utilized protocol for car-diac CT in the individual centers. For ECG-triggered examination, prospective ECG-triggering at 37% of the cardiac cycle was used to allow data acquisition at the lowest velocity (Figure 2). Details about image acquisition and reconstruction parameters are shown in Table 2. Each examination was repeated three times on both CT systems with a small random translocation (approximately 5 mm) and rotation (approximately 5 degrees) of the phantom between each examination. The two CT systems were routinely calibrated.

Table 1 Physical property of artificial calcified lesions of high, medium and low density in large, medium and small size

Size	Density	Concentration, mg/cm^3	Volume, mm^3	Calcium mass, mg
Large	High	796	62.8	50.0
	Medium	401	62.8	25.2
	Low	197	62.8	12.4
Medium	High	796	24.6	19.6
	Medium	401	24.6	9.9
	Low	197	24.6	4.8
Small	High	796	9.1	7.2
	Medium	401	9.1	3.6
	Low	197	9.1	1.8

Quantitative image analysis

The amount of calcium, expressed as Agatston score [12], was quantified by a software package (Aquarius iNtuition, TeraRecon, Foster City, US), for each calcified lesion, using the widely-utilized threshold of 130 HU in nontriggered and ECG-triggered scan [13]. Thresholds from 90 to 160 HU, with steps of 10 HU, were also used in nontriggered scan to analyze the influence of threshold on the agreement of calcium scores between these two techniques. Calcium scores derived from nontriggered thoracic CT examinations with velocities of 0-90 mm/s were considered as index test results, because these veloci-ties were in the common coronary velocity range in a cardiac cycle [8]. Calcium scores for calcified lesions with velocities of 0-30 mm/s derived from ECG-triggered CT examina-tions were considered as reference standard, because these velocities were in the common coronary velocity range of mid-systolic and mid-diastolic coronary motion, in which a usual ECG-triggered acquisition is performed to minimize motion artifacts [8, 14].

Table 2 CT acquisition protocol and image reconstruction parameters

	CT-A (Siemens Sensation 64)		CT-B (Philips Brilliance 64)	
	Nontriggered	ECG-triggered	Nontriggered	ECG-triggered
Acquisition mode	Spiral	Sequential	Spiral	Sequential
Tube voltage	120 kV	120 kV	120 kV	120 kV
Tube charge	20 mAs	120 mAs	20 mAs	120 mAs
Collimation	32×2×0.6 mm	20×1.2 mm	64×0.625 mm	40×0.625 mm
Pitch	1.5	n/a	1.39	n/a
Rotation time	500 ms	330 ms	500 ms	400 ms
Reconstruction mode	Single sector	Single sector	Single sector	Single sector
Reconstruction Kernel	B30f	B30f	B	B
Field of view	250 mm	250 mm	250 mm	250 mm
Slice thickness	3 mm	3 mm	3 mm	2.5 mm
Slick interval	3 mm	3 mm	3 mm	2.5 mm

ECG = electrocardiographically; n/a = not applicable.

Theoretical assessment of blurring

A quantitative method was used to analyze the calcium scores caused by motion blurring. The cardiac motion susceptibility (CMS) index was calculated, which quantifies the susceptibility of the calcium score to motion [15, 16]:

$$CMS = \frac{1}{N-1}\sqrt{\sum_{i=1}^{N}(x_i - x_0)^2}\,\frac{1}{x_0},$$

in which x_o is the calcium score at rest, x_i is the calcium score at a velocity (v). In nontriggered scan, v = 10, 30, 50, 70 and 90 mm/s, respectively. In ECG-triggered scan, v = 10 and 30 mm/s, respectively. And N is the total number of velocities used (N = 6 for nontriggered scan, N = 3 for ECG-triggered scan). The mean CMS was calculated over three repeating scans and two CT scanners. A smaller CMS index represents lower susceptibil-ity of calcium scores to motion.

Statistical analysis

Results were provided as mean ± standard deviation (SD). Sensitivity to detect calcification was calculated as percentage of nonzero calcium score among all the scores of an artificial calcified lesion in the respective velocity range (0-90 mm/s for nontriggered and

0-30 mm/s for ECG-triggered CT). The agreement in detection of calcification between the nontriggered and ECG-triggered CT examination was expressed by kappa value.

Normality of distribution of calcium scores for each calcified lesion at different velocities (0-90 mm/s for nontriggered CT, 0-30 mm/s for ECG-triggered CT) was assessed by one-sample Kolmogorov-Smirnov test. The reliability of calcium scores between nontriggered and ECG-triggered CT examinations, and between CT-A and CT-B was expressed as intra-class correlation coefficient (ICC). Differences in calcium scores between the nontriggered and ECG-triggered CT examination were evaluated by paired-samples t-test. Then, 95% confidence interval (CI) of calcium scores based on ECG-triggered CT was calculated by mean ± 1.96 SD. Outliers of calcium scores on the nontriggered CT examinations beyond the 95% CI of ECG-triggered CT were descriptively analyzed.

A $p < 0.05$ was considered as statistically significant. Statistical analyses were performed using a software package (SPSS 18.0.3, SPSS, Chicago, US).

Results

Detection of calcified lesions

Images of the nine calcified lesions in nontriggered and, to a lesser extent, in ECG-triggered CT were blurred by motion. Increasing blurring was observed when velocity increased (Figure 4). Nontriggered examinations showed overall higher susceptibility of calcium scores to motion, than ECG-triggered mode. The susceptibility difference between nontriggered and ECG-triggered scans was smaller in medium density and me-dium size lesions, than in large-size high-density lesion, small size and low density le-sions (Figure 5).

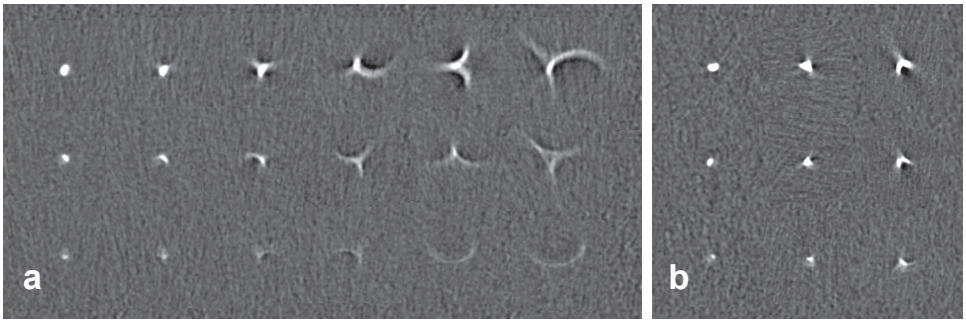


Figure 4 Computed tomography (CT) images of the artificial high-density calcifications in (**a**) the nontriggered thoracic examination (large, medium and small size of the lesion from top to bottom, and velocity of 0, 10, 30, 50, 70 and 90 mm/s from left to right), and (**b**) ECG-triggered cardiac examination (large, medium and small size from top to bottom, and 0, 10 and 30 mm/s from left to right). Window center: 75 HU, window width: 350 HU.

Table 3 Sensitivity to detect coronary calcifications in nontriggered and ECG-triggered CT

Size	Density	Nontriggered							ECG-triggered				*Kappa**
		Sensitivity	Velocity, mm/s						Sensitivity	Velocity, mm/s			
			0	10	30	50	70	90		0	10	30	
Large	High	100%	●	●	●	●	●	●	100%	●	●	●	1.00
	Medium	100%	●	●	●	●	●	●	100%	●	●	●	1.00
	Low	58%	●	●	●	○			100%	●	●	●	0.58
Medium	High	100%	●	●	●	●	●	●	100%	●	●	●	1.00
	Medium	100%	●	●	●	●	●	●	100%	●	●	●	1.00
	Low	42%	●	●	○				67%	●	●		0.47
Small	High	50%	●	●	○	○			67%	●	●		0.50
	Medium	33%	●	●					67%	●	●		0.44
	Low	0%							0%				1.00

**kappa* value is for the agreement to detect coronary calcification between nontriggered and ECG-triggered CT.

Sensitivity to detect coronary calcification was calculated by the percentage of nonzero scores among all the scores of an artificial calcified lesion in the velocity range. Black dot = calcification was detected in CT-A and CT-B; white dot = calcification was detected either in CT-A or CT-B; empty = calcification was not detected.

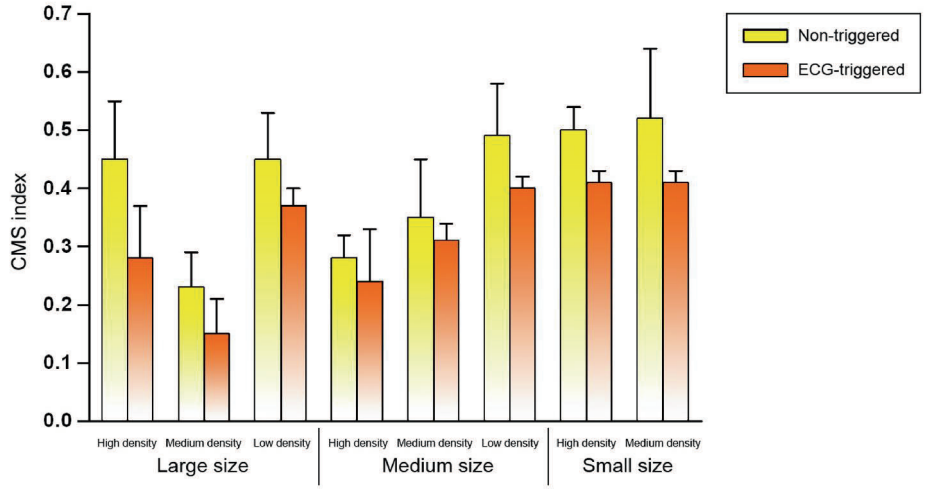


Figure 5 Cardiac motion susceptibility (CMS) index of eight measurable artificial calcified lesions in nontriggered and ECG-triggered CT examinations.

Using CT density threshold of 130 HU in both nontriggered and ECG-triggered examinations, calcium scores of four large calcifications (calcium score > 25 in ECG-triggered mode) could be determined successfully at all velocities with a sensitivity of 100% (Table 3). The agreement to detect large calcifications was excellent (kappa = 1.00) between the nontriggered and ECG-triggered CT. However, as the velocity increased, calcium score assessment failed in four small calcifications (calcium score < 25 in ECG-triggered mode), because the CT attenuation of these lesions dropped under the threshold of 130 HU. For these small calcifications, mean sensitivity of detection was 46 ± 11% in nontriggered CT, and 75 ± 17% in ECG-triggered CT (p value for difference < 0.05). The agreement to detect small calcification was moderate (kappa = 0.44 to 0.58) between the nontriggered and ECG-triggered CT. For the low-density small-size lesion, the calcium score could not be assessed because the CT value was always < 90 HU at all velocities. When measurement failed, calcium score was considered to be zero.

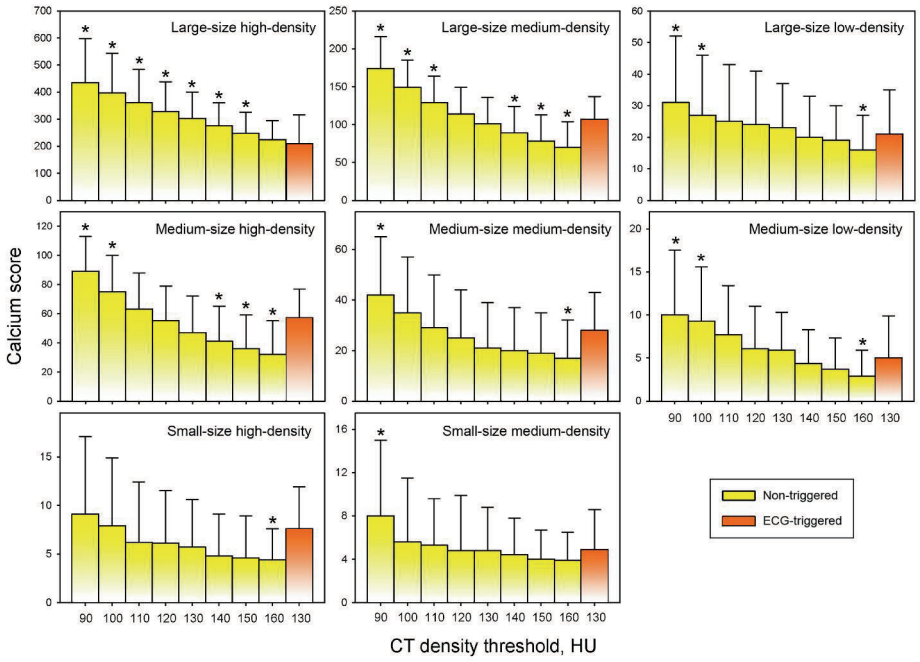


Figure 6 Comparison of calcium scores between nontriggered and ECG-triggered CT examinations in eight measurable calcified lesions. The CT density thresholds were from 90 to 160 HU in nontriggered CT. The threshold was 130 HU in ECG-triggered CT as reference. A significant difference was indicated by "*".

Agreement between nontriggered and ECG-triggered CT

Increasing CT density threshold resulted in decreasing calcium score (Figure 6). The agreement of calcium scores between nontriggered and ECG-triggered scans depended on threshold. E.g., applying thresholds from 90 to 150 HU, significant difference was found for the large-size high-density lesion ($p < 0.05$). When the threshold increased to 160 HU, no significant difference was found ($p > 0.05$). Considering the thresholds, 120 HU and 130 HU showed the highest rate of agreement (7 out of 8 measurable lesions, $p > 0.05$) between nontriggered and ECG-triggered scans, except for the high-density large-size lesion ($p < 0.001$).

Calcium scores in nontriggered and in ECG-triggered CT conformed to a normal distribution (test for normality $p > 0.05$). High reliability of calcium scoring was found between nontriggered and ECG-triggered CT (ICC = 0.960, $p < 0.001$), and between the two CT modalities (ICC = 0.982, $p < 0.001$).

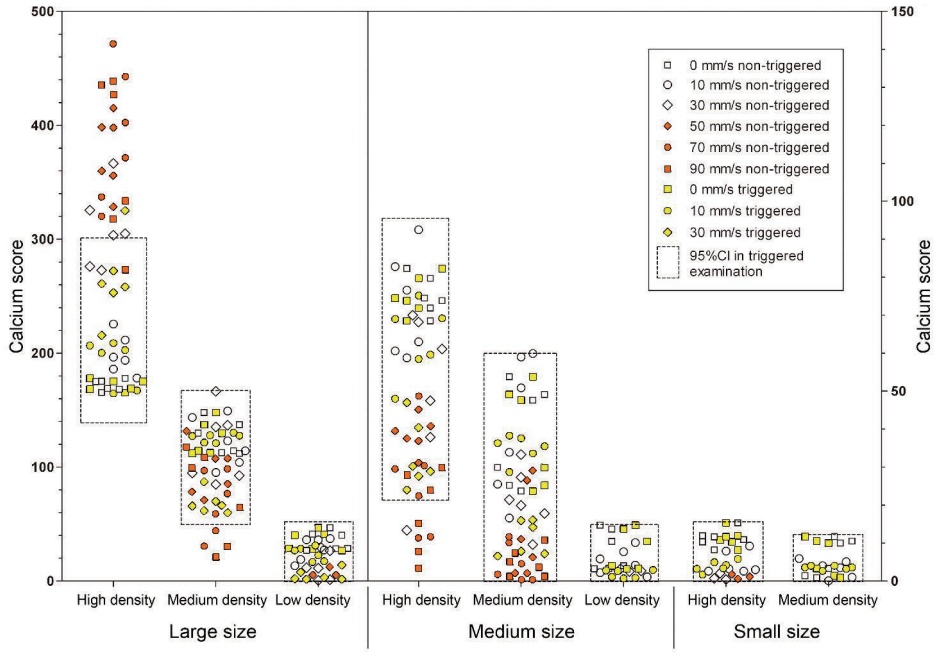


Figure 7 Grouped plots of the calcium scores of the eight measurable artificial calcified lesions in the nontriggered and ECG-triggered CT examinations. The dotted boxes indicate 95% confidence interval (95% CI) of the calcium scores in ECG-triggered CT examinations.

Calcium scores for individual calcifications varied with velocity. Compared to the calcium score at rest, calcium score for high-density and large-size lesions generally showed overestimation at increasing velocity, while calcium score for the other calcifications generally showed underestimation. Most of the calcium scores in nontriggered CT were within the 95%CI of those in ECG-triggered CT. Exceptions to that were the high-density large-size lesion (calcium score 165 - 325 in ECG-triggered mode), mostly when velocity was at least 50 mm/s, the high-density medium-size lesion, and the medium-density large-size lesion, mostly when velocity was at least 70 mm/s (Figure 7).

Discussion

This is the first phantom validation study on the impact of motion artifacts, regarding coronary calcium scoring in nontriggered CT. When performing multiple measurements, good agreement in calcium scores was found between nontriggered low-dose thoracic and ECG-triggered cardiac CT, but not for large-size and high-density calcifications at higher coronary velocity. Nontriggered CT showed lower sensitivity to detect small calcification (calcium score < 25) than ECG-triggered CT.

Good agreement in calcium scores between the nontriggered and ECG-triggered CT was established by distribution analysis, which implies that calcium score derived from low-dose, nontriggered CT can estimate coronary calcium. Coronary calcium reflects the extent of atherosclerotic coronary disease, which is a strong risk marker for cardiac mortality. The relative risk of coronary events increases with increasing coronary calcium burden [4]. In several lung cancer screening trials based on nontriggered thoracic CT, the group of individuals with coronary calcium had higher mortality rate, compared to the group without coronary calcium [17-19]. Compared to mere assessment of presence of coronary calcification, calcium scoring in nontriggered CT may yield information regarding the amount of coronary calcium.

Four in-vivo studies investigated the agreement of calcium scores between nontriggered and ECG-triggered CT [20-23], which included 1,153 subjects in total. In these four studies, 55 subjects (8.8%) showed CS = 0 in nontriggered CT among 625 subjects with CS > 0 in ECG-triggered CT. In concordance with these in-vivo studies, our phantom study validated that a zero calcium score in nontriggered CT does not reliably exclude coronary calcification, thus can not exclude low cardiovascular risk of an individual with a zero CS. On the other hand, nontriggered CT underestimated high-risk CS in 31 (19.1%) subjects, who showed CS < 400 in nontriggered CT, among 162 subjects with CS ≥ 400 in ECG-triggered CT. However, nontriggered CT only overestimated 26 subjects (2.6%), who showed CS≥400 in nontriggered CT, among 991 subjects with CS < 400 in ECG-triggered CT. These results suggest that a high CS detected in nontriggered CT is fairly reliable. In accordance with in-vivo results, our phantom study found that nontriggered CT often underestimated CS, except for the high-density large-size calcified lesion, which is uncommon in clinical practice [24]. From phantom results, we can deduce that a zero CS in nontriggered CT cannot exclude the presence of coronary calcium, and thus can underestimate cardiovascular risk. On the other hand, nontriggered CT can reliably detect a high CS, and thus can likely identify individual at high risk of cardiovascular disease.

Disagreement in calcium scores between nontriggered and ECG-triggered CT examinations was found for the high-density large-size lesion (calcium score 165 to 325). In-vivo, this lesion has not been commonly found, whereas calcified lesions with medium-

and low-density with medium- and small-size are more common [24]. Disagreement was also found for some calcifications at high velocity (≥ 50 mm/s). The velocity of left main artery and left anterior descending artery has been observed to be lower than that of right coronary artery and left circumflex artery [8]. The contribution of left anterior descending artery to calcium score is, on the other hand, higher than that of the other three coronary branches [25]. Deviations in calcium score between nontriggered and ECG-triggered CT were found mainly at high velocities, which do not usually occur in left main and left anterior descending artery [26]. For those reasons, we expect the disagreement in calcium scores between nontriggered and ECG-triggered CT to be less in clinical situations.

In our study, using the calcium scores in ECG-triggered scan as reference, the agreement of calcium scores depended on CT density threshold applied for the nontriggered examination. Using physical lesion properties as reference, Groen et al reported that threshold-adjusted calcium scoring is less susceptible to cardiac motion and more accurate [16]. Thus, a new software based lesion size- and density-dependent threshold might yield better results. In contrast, current software packages commonly utilize a universal threshold. In our results, there was the highest rate of agreement between nontriggered and ECG-triggered results, when using 120 or 130 HU as the threshold, except for the high-density large-size lesion. Also, 130 HU was extensively validated for calcium scoring in coronary risk assessment [4]. Thus, our results confirm the application of 130 HU as a threshold for calcium scoring in current software.

Previous phantom studies have observed less accurate calcium scoring in ECG-triggered 64-MDCT, the CT system used in this study, compared to the reference, electron-beam tomography, as well as dual-source CT [27,28]. Compared to the reference standard, especially underestimated calcium score was found [27]. In the current study, under-estimation by nontriggered 64-MDCT for small calcifications (calcium score < 25 in ECG-triggered mode) appeared even stronger. The sensitivity of nontriggered CT for less dense and smaller calcifications was significantly lower than for ECG-triggered CT (difference in sensitivity up to 42% for individual calcified lesions). The ability to discriminate between a zero and nonzero calcium score is of great importance, as absence of coronary calcification has been found to be associated with an extremely low risk of coronary events in the next years, and is thus a very reassuring finding if accurate [29].

The velocities investigated in this study were in the common motion range of in-vivo coronary arteries. For nontriggered CT examinations, calcium scores at velocities from 0-90 mm/s were applied as index test results because the peak velocity for all segments in coronary artery has been reported to be 88 mm/s [8]. Also, when performing a thoracic low-dose CT, it is unknown at what coronary velocity the individual calcifications are scanned, as data are acquired over entire heart beats. In contrast, ECG-triggered acquisition is performed in the interval of the cardiac cycle with the least motion to reduce

motion artifacts. Velocities have been reported up to 13.0 ± 2.3 mm/s when triggered in mid-systole (for heart rates below 83 bpm), and up to 26.1 ± 2.3 mm/s in mid-diastole (for heart rates above 83 bpm) [8]. This was the basis for our decision to consider calcium scores at velocities of 0-30 mm/s in ECG-triggered CT as reference.

The influence of motion on calcium score in the nontriggered CT examinations strongly depended on velocity. Similar results have been found in ECG-triggered CT for coronary calcium quantification [30-32]. This finding can be explained by blurring in images and partial volume effects [28]. Motion artifacts generally cause overestimation of calcium volume and underestimation of CT density (except in case of smaller calcifications - these are potentially missed), thus leading to inaccurate calcium scores. Individual calcifications in nontriggered CT are likely imaged at different intervals of the cardiac cycle, with different deviations from the optimal calcium score, derived if ECG-triggering would be used. For the individual patient, differences in calcium scores between nontriggered and ECG-triggered acquisition can be positive or negative.

Calcium score according to Agatston was calculated in this study. In a multimanufacturer standard for cardiac CT, mass score was promoted for coronary calcium quantification, as this scoring method showed the lowest variance [33]. However, in clinical practice, so far Agatston score has been mostly utilized to quantify coronary calcium, and to stratify cardiovascular risk. Moreover, in a large screening study, only small differences in cardiovascular risk categorization were found in case of Agatston score versus mass score [34].

In traditional ECG-triggered cardiac CT, often small FOV size is used, in order to improve the detection of coronary calcifications [13]. The standard reconstructed slice thickness is 3 mm. Conversely, in nontriggered thoracic CT often a large FOV (≈ 350 mm) is applied, combined with thin-slices (≈ 1 mm). When evaluating calcium scores in nontriggered mode, an additional dataset should be reconstructed, using a small FOV (≈ 250 mm) and somewhat thicker slices (≈ 3 mm) [22,23,35].

The primary aim of this study was to assess in a controlled environment, whether nontriggered CT, i.e., thoracic CT, can be used to determine the coronary calcium score. The background of this question is the increasing application of thoracic CT, for example in lung cancer screening. In a lung cancer screening trial, over 70% of the participants showed coronary calcifications [19]. The group at risk for lung cancer overlaps with the group at risk for cardiovascular disease, because aging and smoking are two major risk factors for both diseases [36]. There may be substantial primary prevention potential if the calcium score can be derived from the same CT examination. Based on these phantom results, nontriggered CT cannot be promoted as a primary tool to evaluate the calcium score, instead of a dedicated, ECG-triggered cardiac CT. However, the current study does suggest that a nontriggered, thoracic CT examination can additionally yield information

on cardiovascular risk as estimated by the calcium score. From this phantom study, it can already be concluded that a high calcium score can be detected by nontriggered CT, and thus, that nontriggered CT likely can identify individuals at high risk of cardiovascular disease. On the other hand, the calcium score derived from nontriggered CT cannot exclude low risk, as the absence of coronary calcification cannot reliably be determined.

There are limitations to our study. Firstly, the movement of the artificial calcified lesions was linear. The in-vivo coronary arteries exhibit, however, diverse and complex motion patterns in different planes. Because the right coronary artery exhibits the largest motion, and shows the largest artifacts in ECG-triggered scans [26,37], we simulated the motion in the direction of the primary movement of the right coronary artery, in the *x*-*y* plane. It is very difficult to accurately simulate the complex motion of coronary arteries. Further study is needed to confirm our findings in a phantom with multilinear movement, more resembling the actual motion of coronary arteries in three-dimensional space. Secondly, the calcium score could not be assessed for the low-density small-size lesion, because the CT value was always < 90 HU at all velocities. In a previous study using the same phantom with same calcium inserts, this low-density small-size lesion was not detectable or measurable either, even in a stationary setting [38]. Thirdly, calcium score depends on modality, acquisition protocol and scoring parameter setting [15,27]. Our acquisition protocol for nontriggered examinations was derived from the NELSON trial [39], which is a common protocol for low-dose CT lung cancer screening. Finally, because of the setting for CT imaging from different vendors, several parameters were different between CT-A and CT-B, i.e., pitch, rotation time and slice thickness, which could have potentially increased inter-scanner variability.

Conclusions

In this phantom study, when performing multiple measurements, overall good agreement in positive calcium scores was found between nontriggered thoracic and ECG-triggered cardiac CT examinations. However, agreement decreased with increasing coronary velocity. Variability existed among different measurements of a calcified lesion. Nontriggered CT was less sensitive to detect small calcifications than ECG-triggered CT. Our phantom study supports the potential to evaluate coronary calcium score in nontriggered CT. From our results, it can already be concluded that calcium scoring based on nontriggered CT likely can identify individuals at high risk of cardiovascular disease. However, a zero calcium score in nontriggered CT does not reliably exclude the presence of coronary calcification. Further study is needed with calcifications that represent the entire spectrum of coronary calcifications, using a moving phantom with multilinear motion, to confirm and strengthen these results.

References

1. Elias-Smale SE, Proenca RV, Koller MT, Kavousi M, van Rooij FJ, Hunink MG, et al. Coronary calcium score improves classification of coronary heart disease risk in the elderly: the Rotterdam study. J Am Coll Cardiol. 2010;56:1407-14.
2. Nasir K, Rubin J, Blaha MJ, Shaw LJ, Blankstein R, Rivera JJ, et al. Interplay of coronary artery calcification and traditional risk factors for the prediction of all-cause mortality in asymptomatic individuals. Circulation Cardiovascular imaging. 2012;5:467-73.
3. Greenland P, Bonow RO, Brundage BH, Budoff MJ, Eisenberg MJ, Grundy SM, et al. ACCF/AHA 2007 clinical expert consensus document on coronary artery calcium scoring by computed tomography in global cardiovascular risk assessment and in evaluation of patients with chest pain: a report of the American College of Cardiology Foundation Clinical Expert Consensus Task Force (ACCF/AHA Writing Committee to Update the 2000 Expert Consensus Document on Electron Beam Computed Tomography). Circulation. 2007;115:402-26.
4. Oudkerk M, Stillman AE, Halliburton SS, Kalender WA, Mohlenkamp S, McCollough CH, et al. Coronary artery calcium screening: current status and recommendations from the European Society of Cardiac Radiology and North American Society for Cardiovascular Imaging. Eur Radiol. 2008;18:2785-807.
5. Stolzmann P, Leschka S, Betschart T, Desbiolles L, Flohr TG, Marincek B, et al. Radiation dose values for various coronary calcium scoring protocols in dual-source CT. The international journal of cardiovascular imaging. 2009;25:443-51.
6. Berrington de Gonzalez A, Mahesh M, Kim KP, Bhargavan M, Lewis R, Mettler F, et al. Projected cancer risks from computed tomographic scans performed in the United States in 2007. Arch Intern Med. 2009;169:2071-7.
7. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med. 2011;365:395-409.
8. Husmann L, Leschka S, Desbiolles L, Schepis T, Gaemperli O, Seifert B, et al. Coronary artery motion and cardiac phases: dependency on heart rate -- implications for CT image reconstruction. Radiology. 2007;245:567-76.
9. Scott DS, Arora UK, Farb A, Virmani R, Weissman NJ. Pathologic validation of a new method to quantify coronary calcific deposits in vivo using intravascular ultrasound. Am J Cardiol. 2000;85:37-40.
10. Moselewski F, Ferencik M, Achenbach S, Abbara S, Cury RC, Booth SL, et al. Threshold-dependent variability of coronary artery calcification measurements -

implications for contrast-enhanced multi-detector row-computed tomography. Eur J Radiol. 2006;57:390-5.

11. Xie X, Zhao Y, Snijder R, van Ooijen P, Oudkerk M, de Bock G, et al. Sensitivity and accuracy of volumetry of pulmonary nodules on low-dose 16- and 64-row multi-detector CT: an anthropomorphic phantom study. Eur Radiol 2013;23:139-47.
12. Agatston AS, Janowitz WR, Hildner FJ, Zusmer NR, Viamonte M, Jr., Detrano R. Quantification of coronary artery calcium using ultrafast computed tomography. J Am Coll Cardiol. 1990;15:827-32.
13. Mahnken AH, Muhlenbruch G, Koos R, Das M, Pohl S, Stanzel S, et al. Influence of a small field-of-view size on the detection of coronary artery calcifications with MSCT: in vitro and in vivo study. European Radiology. 2006;16:358-64.
14. Greuter MJW, Flohr T, van Ooijen PMA, Oudkerk M. A model for temporal resolution of multidetector computed tomography of coronary arteries in relation to rotation time, heart rate and reconstruction algorithm. European Radiology. 2007;17:784-812.
15. Greuter MJW, Groen JM, Nicolai LJ, Dijkstra H, Oudkerk M. A model for quantitative correction of coronary calcium scores on multidetector, dual source, and electron beam computed tomography for influences of linear motion, calcification density, and temporal resolution: A cardiac phantom study. Medical Physics. 2009;36:5079-88.
16. Groen JM, Dijkstra H, Greuter MJW, Oudkerk M. Threshold adjusted calcium scoring using CT is less susceptible to cardiac motion and more accurate. Medical Physics. 2009;36:438-46.
17. Sverzellati N, Cademartiri F, Bravi F, Martini C, Gira FA, Maffei E, et al. Relationship and Prognostic Value of Modified Coronary Artery Calcium Score, FEV_1, and Emphysema in Lung Cancer Screening Population: The MILD Trial. Radiology. 2012;262:460-7.
18. Shemesh J, Henschke CI, Shaham D, Yip R, Farooqi AO, Cham MD, et al. Ordinal Scoring of Coronary Artery Calcifications on Low-Dose CT Scans of the Chest is Predictive of Death from Cardiovascular Disease. Radiology. 2010;257:541-48.
19. Jacobs PC, Gondrie MJA, van der Graaf Y, de Koning HJ, Isgum I, van Ginneken B, et al. Coronary artery calcium can predict all-cause mortality and cardiovascular events on low-dose CT screening for lung cancer. AJR American journal of roentgenology. 2012;198:505-11.
20. Budoff MJ, Nasir K, Kinney GL, Hokanson JE, Barr RG, Steiner R, et al. Coronary artery and thoracic calcium on noncontrast thoracic CT scans: comparison of ungated and gated examinations in patients from the COPD Gene cohort. J Cardiovasc Comput Tomogr. 2011;5:113-8.

21. Einstein AJ, Johnson LL, Bokhari S, Son J, Thompson RC, Bateman TM, et al. Agreement of Visual Estimation of Coronary Artery Calcium From Low-Dose CT Attenuation Correction Scans in Hybrid PET/CT and SPECT/CT With Standard Agatston Score. J Am Coll Cardiol. 2010;56:1914-21.
22. Kim SM, Chung MJ, Lee KS, Choe YH, Yi CA, Choe BK. Coronary calcium screening using low-dose lung cancer screening: effectiveness of MDCT with retrospective reconstruction. AJR American journal of roentgenology. 2008;190:917-22.
23. Wu MT, Yang P, Huang YL, Chen JS, Chuo CC, Yeh C, et al. Coronary arterial calcification on low-dose ungated MDCT for lung cancer screening: concordance study with dedicated cardiac CT. AJR Am J Roentgenol. 2008;190:923-8.
24. Qian Z, Anderson H, Marvasty I, Akram K, Vazquez G, Rinehart S, et al. Lesion- and vessel-specific coronary artery calcium scores are superior to whole-heart Agatston and volume scores in the diagnosis of obstructive coronary artery disease. J Cardiovasc Comput Tomogr. 2010;4:391-9.
25. Schmermund A, Baumgart D, Mohlenkamp S, Kriener P, Pump H, Gronemeyer D, et al. Natural history and topographic pattern of progression of coronary calcification in symptomatic patients: An electron-beam CT study. Arterioscler Thromb Vasc Biol. 2001;21:421-6.
26. Achenbach S, Ropers D, Holle J, Muschiol G, Daniel WG, Moshage W. In-plane coronary arterial motion velocity: measurement with electron-beam CT. Radiology. 2000;216:457-63.
27. Greuter MJ, Dijkstra H, Groen JM, Vliegenthart R, de Lange F, Renema WK, et al. 64 slice MDCT generally underestimates coronary calcium scores as compared to EBT: a phantom study. Med Phys. 2007;34:3510-9.
28. Groen JM, Greuter MJ, Vliegenthart R, Suess C, Schmidt B, Zijlstra F, et al. Calcium scoring using 64-slice MDCT, dual source CT and EBT: a comparative phantom study. Int J Cardiovasc Imaging. 2008;24:547-56.
29. Sarwar A, Shaw LJ, Shapiro MD, Blankstein R, Hoffmann U, Cury RC, et al. Diagnostic and prognostic value of absence of coronary artery calcification. JACC Cardiovascular imaging. 2009;2:675-88.
30. Groen JM, Greuter MJ, Schmidt B, Suess C, Vliegenthart R, Oudkerk M. The influence of heart rate, slice thickness, and calcification density on calcium scores using 64-slice multidetector computed tomography - A systematic phantom study. Investigative Radiology. 2007;42:848-55.
31. King M, Giger ML, Suzuki K, Bardo DME, Greenberg B, Lan L, et al. Computerized assessment of motion-contaminated calcified plaques in cardiac multidetector CT. Medical Physics. 2007;34:4876-89.

32. King M, Giger ML, Suzuki K, Pan XC. Feature-based characterization of motion-contaminated calcified plaques in cardiac multidetector CT. Medical Physics. 2007;34:4860-75.
33. McCollough CH, Ulzheimer S, Halliburton SS, Shanneik K, White RD, Kalender WA, et al. Coronary artery calcium: A multi-institutional, multimanufacturer international standard for quantification at cardiac CT. Radiology. 2007;243:527-38.
34. Rumberger JA, Kaufman L. A rosetta stone for coronary calcium risk stratification: agatston, volume, and mass scores in 11,490 individuals. AJR Am J Roentgenol. 2003;181:743-8.
35. Isgum I, Prokop M, Niemeijer M, Viergever MA, van Ginneken B. Automatic coronary calcium scoring in low-dose chest computed tomography. IEEE transactions on medical imaging. 2012;31:2322-34.
36. Oei HH, Vliegenthart R, Hofman A, Oudkerk M, Witteman JC. Risk factors for coronary calcification in older subjects. The Rotterdam Coronary Calcification Study. European heart journal. 2004;25:48-55.
37. Ohnesorge B, Flohr T, Fischbach R, Kopp AF, Knez A, Schroder S, et al. Reproducibility of coronary calcium quantification in repeat examinations with retrospectively ECG-gated multisection spiral CT. Eur Radiol. 2002;12:1532-40.
38. Vliegenthart R, Song B, Hofman A, Witteman JC, Oudkerk M. Coronary calcification at electron-beam CT: effect of section thickness on calcium scoring in vitro and in vivo. Radiology. 2003;229:520-5.
39. Ru Zhao Y, Xie X, de Koning HJ, Mali WP, Vliegenthart R, Oudkerk M. NELSON lung cancer screening study. Cancer imaging. 2011;11 S79-84.

Chapter 8

Morphological Measurements in Computed Tomography Correlate with Airflow Obstruction in Chronic Obstructive Pulmonary Disease: Systematic Review and Meta-Analysis

European Radiology. 2012; 22: 2085-2093

Xueqian Xie[12] / Pim A. de Jong[3] / Matthijs Oudkerk[2] / Ying Wang[4]
Nick H.T. ten Hacken[5] / Jingtao Miao[6] / Guixiang Zhang[6]
Geertruida H. de Bock, PhD[7] / Rozemarijn Vliegenthart, MD, PhD[12]

University of Groningen, University Medical Center Groningen,
[1]Department of Radiology, [2]Center for Medical Imaging – North East Netherlands,
[5]Department of Pulmonary Diseases, [7]Department of Epidemiology;
[3]Department of Radiology, University Medical Center Utrecht;
[4]Department of Radiology, Tianjin Medical University General Hospital;
[6]Department of Radiology, Shanghai Jiaotong University Affiliated First People's Hospital

Abstract

Objective

To determine the correlation between CT measurements of emphysema or peripheral airways and airflow obstruction in chronic obstructive pulmonary disease (COPD).

Materials and Methods

PubMed, Embase and Web of Knowledge were searched from 1976 to 2011. Two reviewers independently screened 1,763 citations to identify articles that correlated CT measurements to airflow obstruction parameters of pulmonary function test in COPD patients, rated study quality and extracted information. Three CT measurements were accessed: lung attenuation area percentage < -950 Hounsfield units, mean lung density and airway wall area percentage. Two airflow obstruction parameters were accessed: forced expiratory volume in the first second as percentage from predicted (FEV_1 %pred) and FEV_1 divided by the forced volume vital capacity.

Results

79 articles (9,559 participants) were included in systematic review, demonstrating different methodologies, measurements and CT-airflow obstruction correlations, 15 high-quality articles (2,095 participants) in the meta-analysis. The absolute pooled correlation coefficients ranged from 0.48 (95% CI, 0.40 to 0.54) to 0.65 (0.58 to 0.71) for inspiratory CT and 0.64 (0.53 to 0.72) to 0.73 (0.63 to 0.80) for expiratory CT.

Conclusions

CT measurements of emphysema or peripheral airways are significantly related to airflow obstruction in COPD patients. CT provides a morphological method to investigate airway obstruction in COPD.

Introduction

Chronic obstructive pulmonary disease (COPD) is characterized by airflow limitation that is not fully reversible [1]. The pathogenesis of airflow limitation in COPD is mainly related to emphysema and small airway remodelling [2]. Although airflow obstruction parameters in pulmonary function test (PFT) by spirometry are essential in COPD diagnosis, these parameters fail to quantify the proportionate impact of emphysema and small airways disease individually. Morphologic changes can be characterized and quantified by computed tomography (CT), especially by multi-detector CT [3]. For COPD patients, quantitative chest CT are important to understand pathogenesis and the effect of therapeutic interventions [4], and can help to identify those most at risk for acute exacerbations [5]. A long history of CT emphysema quantification exists since the introduction of the 'density-mask' in 1988 [6-8]. Multi-detector CT can accurately evaluate emphysema [9]. However, quantification of air-way remodelling by CT is challenging because of its spatial resolution. Airway wall quantification started over a decade ago, mainly for large airways [3], but investigators have measured peripheral airways down to 0.5 mm lumen diameter [10] and 2.8 mm outer diameter [11]. Measurement of narrowing of CT-detectable airways may estimate the degree of small airway disease [10].

The assumption is that emphysema and peripheral airway wall thickness, as detected by CT, are correlated to airflow obstruction in COPD patients. Study results have been variable and sometimes conflicting. However, some individual studies have been small and underpowered [2, 12-15]. Therefore, we conducted a systemic review and meta-analysis to determine the correlation between emphysema or peripheral airway measurements on inspiratory and expiratory CT, and airflow obstruction in COPD.

Materials and Methods

This study was conducted according to Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) [16].

Data sources and searches

We searched PubMed, Embase, and Web of Knowledge from January, 1976 to December, 2011, from the start of whole-body CT, using terms related to computed tomography and PFT (i.e., lung function*, respiratory function*, pulmonary function*, etc) and COPD (i.e., chronic obstructive pulmonary disease*, chronic obstructive lung disease*, etc) without language restrictions (Supplemental Table A). Unpublished studies were not included.

Study selection

Four reviewers with at least six years experience in thoracic radiology participated in study selection. Each study was evaluated independently by two reviewers out of three, with disagreements resolved by the fourth reviewer. Articles were included in the systematic review if they: 1) analyzed the association between CT quantitative emphysema or airway measurements and PFT; 2) investigated human beings; 3) included participants diagnosed with stable adult COPD, according to the Global initiative for chronic Obstructive Lung Disease (GOLD) [1] or the American Thoracic Society (ATS) or the European Respiratory Society (ERS) or clearly defined similar criteria; 4) included participants who had clearly described PFT, according to the guideline of ATS or ERS or similar methods. Articles were excluded if they: 1) were reviews or abstracts or case reports or letters; 2) were laboratory or phantom studies; 3) covered participants with confounding disease, such as interstitial lung disease, chronic bronchitis, asthma and α-1 antitrypsin disease.

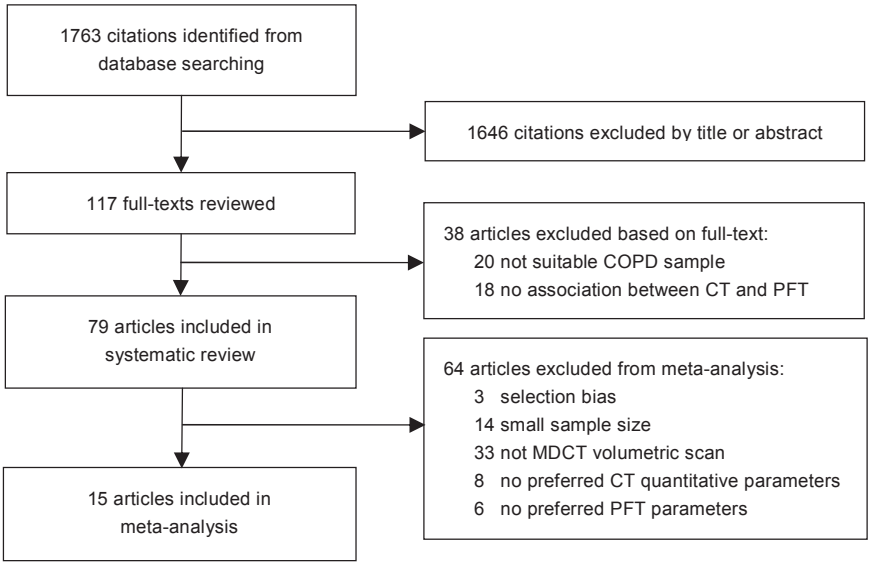


Figure 1 Flowchart of literature review and selection. COPD = chronic obstruction pulmonary disease; PFT = pulmonary function test; MDCT = multi-detector computed tomography.

Articles were subsequently included in the meta-analysis if they: 1) had no selection bias (e.g., only mild or only severe COPD); 2) had a sample size of ≥ 20 (20 subjects would provide a power of 0.90 when detecting a typical effect correlation coefficient of 0.60); 3) were performed using volumetric multi-detector CT; 4) reported correlation coefficients; 5) reported percentage of lung attenuation area under -950 HU (%LAA-950) or mean lung density (MLD) or wall area percentage (WA%) in airways ≥ 5^{th} airway generation (sub-sub-

segment level) as CT measurements; 6) reported the predicted forced expiratory volume in the first second as percentage (FEV_1%pred) and FEV_1 divided by the forced volume vital capacity (FEV_1/FVC) as spirometry parameters. Studies were excluded if the CT examination only included selected pulmonary levels or if slice increment was larger than slice thickness. In possible duplicate reports, the report with largest sample size was included.

Data extraction and quality assessment

Two reviewers evaluated independently, with disagreements resolved by another reviewer. A standardized extraction form was used to collect study characteristics, participant characteristics, methodology and correlation coefficients. The systematic review included 10 CT measurements: %LAA-960, %LAA-950, %LAA-910, %LAA-900, MLD, 15 percentile point of lung density (Perc 15), lung volume (LV), WA%, airway wall thickness (WT) and airway lumen area (Ai). Three CT measurements, %LAA-950, MLD and WA% were pooled in the meta-analysis. Two PFT parameters for airflow obstruction were collected, including FEV_1 %pred and FEV_1/FVC.

Methodological quality and potential sources of bias of the included meta-analysis articles were assessed with 14 standard items of the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool [17]. For each article, a quality score was accumulated by assigning 1 point to each fulfilled QUADAS item, 0.5 to unclear items, and 0 to unmet items. A score of ≥ 11 points was considered high quality and a score < 11 points as low quality. Cohen's *k* was calculated to indicate inter-observer agreement. Publication bias was evaluated with Begg and Mazumdar rank correlation, Egger's regression test and Rosenthal fail-safe N test.

Data synthesis and analysis

Summary measure was the correlation coefficient (CC). Pooled CCs with 95% confidence intervals (CIs) were calculated using Hedges-Vevea random effects model and Z-test for normality. Pooled CCs were calculated for the correlations between %LAA-950 and FEV_1 %pred, %LAA-950 and FEV_1/FVC, MLD and FEV_1 %pred, MLD and FEV_1/FVC, WA% and FEV_1 %pred in inspiratory and expiratory CT. If multi-level bronchi were evaluated, we chose the smallest bronchi. Heterogeneity was tested using Q statistic and I^2 index. The random effects model was used regardless of the heterogeneity test, although results in Q statistic were still stated. To investigate the impact of individual variables on the meta-analysis results, subgroup analysis was performed if a subgroup consisted of at least two studies. Subgroups were based on radiation dose (low or normal dose), and breath-hold procedure (inspiratory or expiratory). Meta-regression was performed to investigate the influence from gender, if male percentage was reported by at least three studies. Statistical

analysis was performed using SPSS 18.0 (IBM, New York, USA) and R 2.12.0 (R Foundation, Vienna, Austria). A $p < 0.05$ was considered statistically significant.

Results

Study selection

The database searches elicited 1,763 citations (Figure 1). 79 articles were included in the systematic review and fifteen articles [13, 18-31] in the meta-analysis, including ten [18-23, 25-26, 28-29] from trial cohorts.

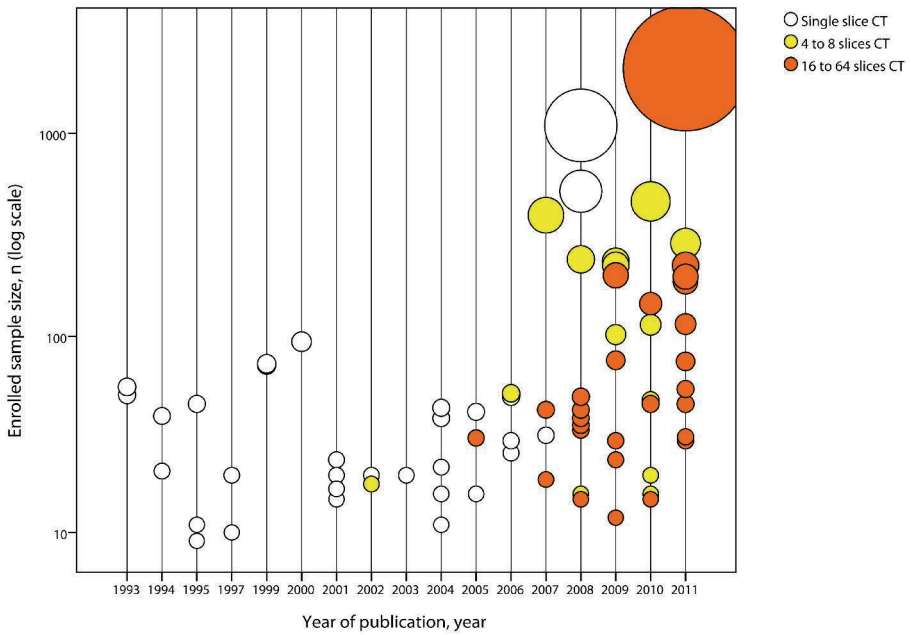


Figure 2 Sample size of the articles included in the systematic review, by year of publication and CT generation. MDCT = multi-detector computed tomography.

Systematic review

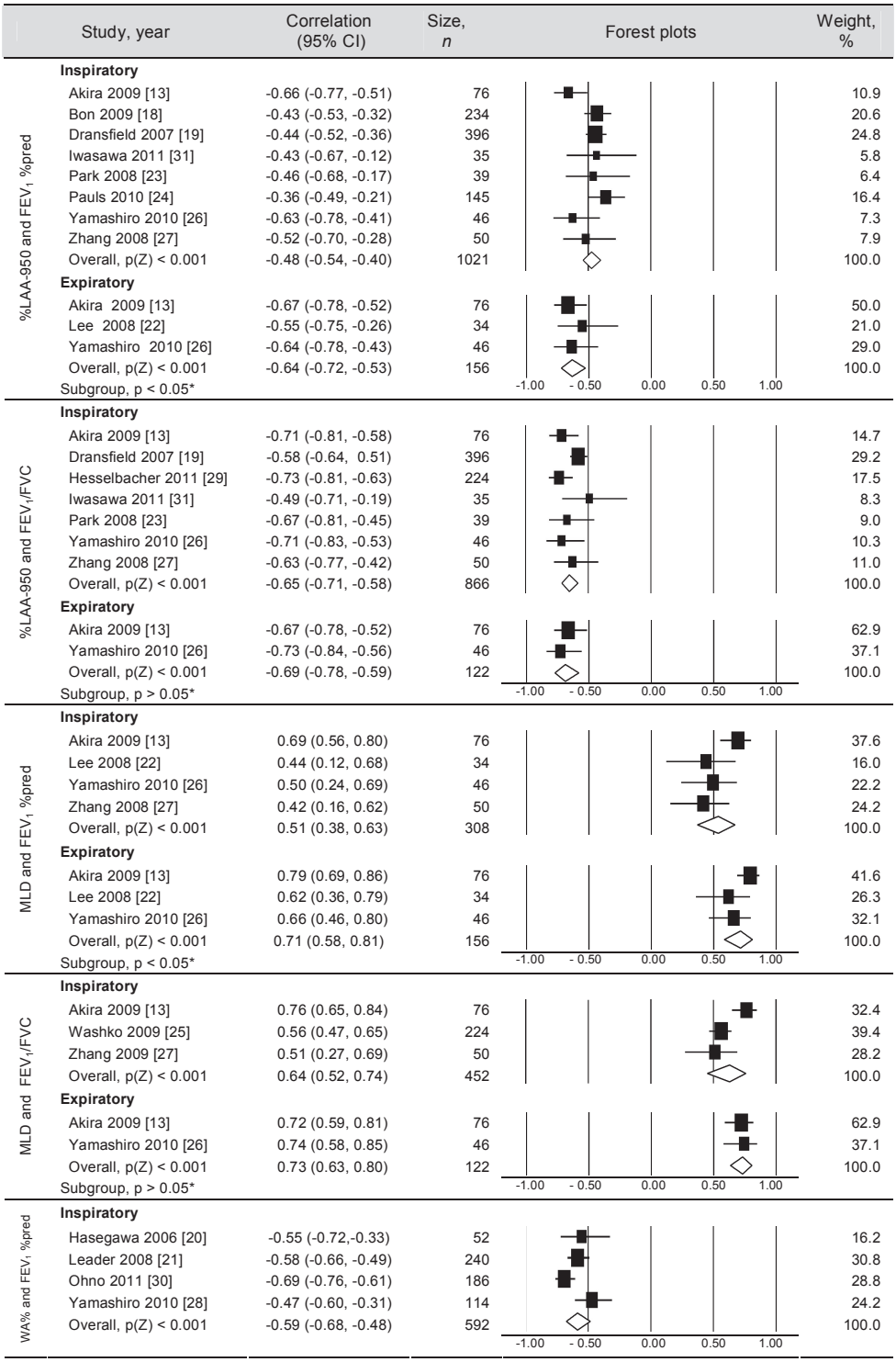
The systematic review included 9,559 COPD participants (range of mean age, 48 to 73 years), including 6,101 (63.8%) men, 2,000 (20.9%) women and 1,458 (15.3%) without indicated gender (Supplemental Table B). 6,935 (72.5%) were (ex-) smokers, 133 (1.4%) non-smokers, and in 2,491 (26.1%) no smoking status was reported. Fifty-four (68.4%) studies were prospective, 23 (29.1%) retrospective, and in 2 (2.5%) articles the study design was not reported. Forty-one (51.9%) articles were from Europe, 23 (29.1%) from Asia and 15 (19.0%)

from North America. Of the articles, 69 (87.3%) were written in English, 5 (6.3%) in Italian, 3 (3.8%) in Chinese, 1 (1.3%) in French and 1 (1.3%) in Polish.

Tendency was towards larger sample size and more advanced CT equipment in recent publications (Figure 2). Before 2007, only articles with sample size < 100 were found. After 2007, larger sample size articles were published. Although single-slice CT (32 articles, 40.5%) was continuously used from 1993 to 2009, multi-detector CT was popular in recent years. From 2002, 4- and 8-slice multi-detector CT (16 articles, 20.2%) was used. From 2005, 16- and 64-slice multi-detector CT (29 articles, 36.7%) was used.

Included articles varied in methodology. Volume acquisition was used in 36 articles (45.6%), nonvolume acquisition in 43 (54.4%). Selected slices were scanned in 21 studies (26.6%), whole lung in 58 (73.4%). Median slice thickness was 1.0 mm (range, 0.625 to 10 mm). Median slice increment was 2.0 mm (range, 0.625 to 20 mm). Low radiation dose was used in 12 articles (15.2%), normal dose in 67 (84.7%). Both inspiratory and expiratory CT findings were evaluated in 24 articles (30.4%), only inspiratory in 54 articles (68.3%) and expiratory in one article (1.3%).

Sixty-four different CT measurements and 27 different PFT parameters were reported (Supplemental Figures A and B). Common CT measurements were %LAA-950 in 36 (45.6%) studies, MLD in 22 (27.8%), WA% in 17 (21.5%) and visual score in 17 (21.5%). Common PFT parameters were FEV_1 %pred in 72 articles (91.1%) and FEV_1/FVC in 64 articles (81.0%). Common lung parenchyma thresholds defining emphysema ranged from -900 HU to -960 HU, the most commonly used threshold being -950 HU. In some studies, different correlations to airflow obstruction parameters in PFT were found with these differing thresholds in the same sample [12, 14, 27, 32]. The CC between %LAA-950 and FEV_1 %pred ranged from -0.67 to -0.09 [12-13], between MLD and FEV_1 %pred from 0.18 to 0.85 [12, 33], between WA% and FEV_1 %pred from -0.713 to -0.044 [22, 34], between %LAA-950 and FEV_1/FVC -0.75 to -0.09 [12, 35], and between MLD and FEV_1/FVC from 0.21 to 0.89 [12, 33]. In four articles, the CC between Perc 15 and FEV_1 %pred in inspiration ranged from 0.09 to 0.62, and the CC between Perc 15 and FEV_1/FVC in inspiration from 0.12 to 0.62 [12-13, 36-37] (Supplemental Table C).


Study, year
Correlation (95% CI)
Size, n
Forest plots
Weight, %
%LAA-950 and FEV1 %pred
Inspiratory
Akira 2009 [13] -0.66 (-0.77, -0.51) 76 10.9
Bon 2009 [18] -0.43 (-0.53, -0.32) 234 20.6
Dransfield 2007 [19] -0.44 (-0.52, -0.36) 396 24.8
Iwasawa 2011 [31] -0.43 (-0.67, -0.12) 35 5.8
Park 2008 [23] -0.46 (-0.68, -0.17) 39 6.4
Pauls 2010 [24] -0.36 (-0.49, -0.21) 145 16.4
Yamashiro 2010 [26] -0.63 (-0.78, -0.41) 46 7.3
Zhang 2008 [27] -0.52 (-0.70, -0.28) 50 7.9
Overall, p(Z) < 0.001 -0.48 (-0.54, -0.40) 1021 100.0
Expiratory
Akira 2009 [13] -0.67 (-0.78, -0.52) 76 50.0
Lee 2008 [22] -0.55 (-0.75, -0.26) 34 21.0
Yamashiro 2010 [26] -0.64 (-0.78, -0.43) 46 29.0
Overall, p(Z) < 0.001 -0.64 (-0.72, -0.53) 156 100.0
Subgroup, p < 0.05*
%LAA-950 and FEV1/FVC
Inspiratory
Akira 2009 [13] -0.71 (-0.81, -0.58) 76 14.7
Dransfield 2007 [19] -0.58 (-0.64, 0.51) 396 29.2
Hesselbacher 2011 [29] -0.73 (-0.81, -0.63) 224 17.5
Iwasawa 2011 [31] -0.49 (-0.71, -0.19) 35 8.3
Park 2008 [23] -0.67 (-0.81, -0.45) 39 9.0
Yamashiro 2010 [26] -0.71 (-0.83, -0.53) 46 10.3
Zhang 2008 [27] -0.63 (-0.77, -0.42) 50 11.0
Overall, p(Z) < 0.001 -0.65 (-0.71, -0.58) 866 100.0
Expiratory
Akira 2009 [13] -0.67 (-0.78, -0.52) 76 62.9
Yamashiro 2010 [26] -0.73 (-0.84, -0.56) 46 37.1
Overall, p(Z) < 0.001 -0.69 (-0.78, -0.59) 122 100.0
Subgroup, p > 0.05*
MLD and FEV1 %pred
Inspiratory
Akira 2009 [13] 0.69 (0.56, 0.80) 76 37.6
Lee 2008 [22] 0.44 (0.12, 0.68) 34 16.0
Yamashiro 2010 [26] 0.50 (0.24, 0.69) 46 22.2
Zhang 2008 [27] 0.42 (0.16, 0.62) 50 24.2
Overall, p(Z) < 0.001 0.51 (0.38, 0.63) 308 100.0
Expiratory
Akira 2009 [13] 0.79 (0.69, 0.86) 76 41.6
Lee 2008 [22] 0.62 (0.36, 0.79) 34 26.3
Yamashiro 2010 [26] 0.66 (0.46, 0.80) 46 32.1
Overall, p(Z) < 0.001 0.71 (0.58, 0.81) 156 100.0
Subgroup, p < 0.05*
MLD and FEV1/FVC
Inspiratory
Akira 2009 [13] 0.76 (0.65, 0.84) 76 32.4
Washko 2009 [25] 0.56 (0.47, 0.65) 224 39.4
Zhang 2009 [27] 0.51 (0.27, 0.69) 50 28.2
Overall, p(Z) < 0.001 0.64 (0.52, 0.74) 452 100.0
Expiratory
Akira 2009 [13] 0.72 (0.59, 0.81) 76 62.9
Yamashiro 2010 [26] 0.74 (0.58, 0.85) 46 37.1
Overall, p(Z) < 0.001 0.73 (0.63, 0.80) 122 100.0
Subgroup, p > 0.05*
WA% and FEV1 %pred
Inspiratory
Hasegawa 2006 [20] -0.55 (-0.72,-0.33) 52 16.2
Leader 2008 [21] -0.58 (-0.66, -0.49) 240 30.8
Ohno 2011 [30] -0.69 (-0.76, -0.61) 186 28.8
Yamashiro 2010 [28] -0.47 (-0.60, -0.31) 114 24.2
Overall, p(Z) < 0.001 -0.59 (-0.68, -0.48) 592 100.0
-1.00 - 0.50 0.00 0.50 1.00

Risk of bias in the meta-analysis

All articles included in the meta-analysis were high quality (Supplemental Table D and Supplemental Figure C). The quality score ranged from 12.5 to 13.5. Suboptimal scores were present for three QUADAS items: seven articles without interval between CT and PFT (item 4), nine articles without indication whether CT quantification was blinded to PFT (item 10), and thirteen articles without indication whether PFT was blinded to CT quantification (item 11). Cohen's *k* was 0.925, expressing very good inter-observer agree-ment. No publication bias was found (Supplemental Table E). The median of Rosenthal fail-safe N was 122 (range, 84 to 614), indicating a solid empirical result.

Three of the nine meta-analysis calculations showed mild heterogeneity. I^2 index was > 50% for the correlation between MLD and FEV_1 %pred in inspiration (p = 0.11, I^2 index = 50.6%), between MLD and FEV_1/FVC in inspiration (p = 0.02, I^2 index = 75.9%) and between WA% and FEV_1 %pred (p = 0.04, I^2 index = 64.7%).

Synthesis of results in the meta-analysis

The meta-analysis included 2,095 participants out of 9,559 in the systematic review (Figure 3).

Nine articles [13, 18-19, 22-24, 26-28, 31] reported CC between %LAA-950 and FEV_1 %pred in inspiration. Two [19, 25] were from the National Lung Screening Trial (NLST) cohort and two [22-23] from the Korean Obstructive Lung Disease (KOLD) cohort. Due to possible duplicate reporting, articles [22, 25] with smaller sample size were excluded from each cohort. The pooled CC between %LAA-950 and FEV_1 %pred was -0.48 (95%CI: -0.54, -0.40) in inspiration and -0.64 (-0.72, -0.53) in expiration. The pooled CC between %LAA -950 and FEV_1/FVC was -0.65 (-0.71, -0.58) in inspiration and -0.69 (-0.78, -0.59) in expiration.

No potential duplicate report was found for MLD. The pooled CC between MLD and FEV_1 %pred was 0.51 (95%CI: 0.38, 0.63) in inspiration and 0.71 (0.58, 0.81) in expiration. The pooled CC between MLD and FEV_1/FVC was 0.64 (0.52, 0.74) in inspiration and 0.73 (0.63, 0.80) in expiration.

◄ **Figure 3** Forest plots for correlations between CT measurements and airflow obstruction. CI = confidence interval; p(Z) = p value of Z test; FEV_1 %pred = percentage of the predicted forced expiratory volume in the first second; FEV_1/FVC = FEV_1 divided by forced vital capacity; %LAA-950 = percentage lower attenuation area than -950 HU; MLD = mean lung density; Perc 15 = 15 percentile point of lung density; WA% = wall area percentage.

Seven articles [18, 20-22, 25, 28, 30] reported CC between WA% and FEV_1 %pred in inspiration. Two articles [18, 22] were excluded because airway measurements concerned only airways above the 5th airway generation. Another article [25] was excluded because it did not report which airways were measured. Four articles [20-21, 28, 30] were finally included. The lumen diameter of peripheral airways was about 2 to 3 mm in the included articles. The pooled CC between WA% and FEV_1 %pred was -0.59 (95%CI: -0.68, -0.48) in inspiration. Expiratory CT was not used for airway measurements.

Subgroup analysis and meta-regression

Subgroup analysis for radiation dose was performed for the association between %LAA-950 and FEV_1 %pred in inspiration, indicating no significant difference ($p > 0.05$). In low dose [18-19], the pooled CC was -0.44 (95%CI: -0.50, -0.37). In normal dose [13, 23-24, 26-27, 31], the pooled CC was -0.50 (-0.57, -0.42). Subgroup analysis was also performed for inspiratory and expiratory CT examination. A significantly stronger negative correlation was found between %LAA-950 and FEV_1 %pred in expiratory CT ($p < 0.05$), and a stronger positive correlation between MLD and FEV_1 %pred ($p < 0.001$), but no difference was found in the association between %LAA-950 and FEV_1/FVC ($p > 0.05$), or MLD and FEV_1/FVC ($p > 0.05$). In meta-regression for gender contribution, no statistically significant effect modification was found for male percentage ($p > 0.05$) (Supplemental Table E).

Discussion

In this meta-analysis, significant correlations were found between CT measurements of emphysema or peripheral airway and airflow obstruction parameters in PFT in COPD patients, both in inspiratory and expiratory CT examination. The range of absolute correlation coefficients between included CT measurements and airflow obstruction was 0.48 to 0.65 for inspiratory CT and 0.64 to 0.73 for expiratory CT. These results confirm correlations between morphology and function in COPD patients. The confidence in these findings is strong, as results were based on high methodological quality studies without publication bias. Thus, CT provides a quantitative morphological method to investigate the principle components of airway obstruction in COPD, with similar strength of associations with airflow obstruction for CT measurements of emphysema and peripheral airways. The strongest association was found between CT emphysema measurements and FEV_1/FVC, especially in expiratory CT. Our systematic review demonstrated differing methodologies for CT quantification, and contrasting correlations with airflow obstruction in COPD patients.

CT quantification reflects pathophysiological changes in COPD to some degree. The pathological findings of airway limitation are in airways < 2 mm in internal diameter [38].

Such small airways can hardly be measured directly by CT due to spatial resolution limit. However, peripheral airway (≥ 5th generation) wall thickness can be measured as WA%. Destruction of the lung parenchyma (emphysema) can be measured as %LAA-950 or MLD. The morphological contribution from these two pathologic processes is difficult to distinguish by spirometry, but is important for COPD research. The morphological information in CT quantification of the relative predominance of peripheral airway wall disease or emphysema may in the future allow more focused treatment of the predominating COPD phenotype.

This systematic review incorporated ten different CT measurements. Although visual score was common in earlier publications, we did not discuss it, because of its subjective nature. Perc 15 seems an effective measurement for emphysema, but Perc 15 results could not be pooled because of insufficient study number. Only three (%LAA-950, MLD and WA%) of the ten measurements were eventually pooled because study number was sufficient to perform meta-analysis. We investigated two PFT parameters (FEV_1 %pred and FEV_1/FVC) as they are the two most commonly used functional parameters regarding airway limitation.

In subgroup analysis, associations for inspiratory and expiratory CT findings were compared. Some authors have reported that CT measurements in expiration are more closely correlated with airflow obstruction than in inspiration [14, 39-41]. Our results indicate that CT measurements in expiration rather than in inspiration were more correlated with FEV_1 %pred, not with FEV_1/FVC. Whether an additional expiratory CT data acquisition should be performed for COPD evaluation, is debatable. Expiratory CT exposes patients to additional radiation, however with developments in CT technique, the additional dose will likely decrease. Also, we found low radiation dose did not change correlations between CT emphysema quantification and airflow obstruction compared to normal dose. Low dose CT can decrease the overall radiation dose for CT quantitative emphysema evaluation without loss in diagnostic value.

Multiple airway generations were included in the systematic review but only peripheral airways (≥ 5th generation) in the meta-analysis. Some authors investigated airways from the 3rd to 5th or 6th generation, and found that the association between airway wall measurements and PFT was stronger for higher generations than lower generations [20, 28]. Therefore, we only pooled results for airways ≥ 5th generation. Some authors reported moderate associations with larger airways, ranging from -0.39 to -0.54 [18, 42]. In our meta-analysis, the association between wall area percentage of peripheral airways and FEV_1 %pred was -0.59. One factor to keep in mind is the overestimation of airway wall thickness, showing a relative increase with each airway generation [3]. Despite that factor, based on our pooled results, the association between disease of the more peripheral airways (≥ 5th generation) and lung function appears stronger than for lower generation airways (< 5th

generation), suggesting that airway wall thickness measurements on CT should be performed on the smallest airways visible.

This study confirms significant correlations between CT measurements and airflow obstruction in COPD. The correlations were in agreement with some expert narrative reviews [9-10] and individual studies [14-15, 39, 42-43]. Nevertheless, other studies reported weaker associations, like the National Emphysema Treatment Trial (NETT) study [12] and the International COPD Genetics Network (ICGN) study [2]. In NETT and ICGN, predominantly single-slice CT was used. Since single-slice CT decreases reproducibility and accuracy [44], that has likely caused the reduced strength of the correlations.

This study has some limitations. First, no prospective large cohort with up-to-date CT technology was found as primary study. The largest study [45] included over 2,000 participants, but most of the participants had normal lung function. In inspiratory CT, the number of included articles was relatively small but sufficient to perform a reliable meta-analysis. However, in expiratory CT, some indicators of bias could not be determined due to the limited number of studies. Second, we found 64 different quantitative CT measurements in the literature. We chose ten for systematic review and three for meta-analysis. Although included parameters are representative, the other 54 measurements could be valuable to evaluate COPD. In another way, FEV_1 %pred and FEV_1/FVC were selected as airflow obstruction parameters in PFT, but other parameters in PFT were valuable to evaluate COPD. Third, mild heterogeneity was found in three correlations in the meta-analysis. A random effects model was used to compensate for the heterogeneity.

In conclusion, measurements of emphysema and peripheral airways on inspiratory and expiratory CT have significant correlations with airflow obstruction as accessed by FEV_1 %pred and FEV_1/FVC in COPD patients. Thus, CT provides a quantitative morphological method to investigate airflow obstruction by emphysema and peripheral airway disease in COPD.

References

1. Rabe KF, Hurd S, Anzueto A, et al. (2007) Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease - GOLD executive summary. American Journal of Respiratory and Critical Care Medicine 176:532-555.
2. Patel BD, Coxson HO, Pillai SG, et al. (2008) Airway wall thickening and emphysema show independent familial aggregation in chronic obstructive pulmonary disease. American Journal of Respiratory and Critical Care Medicine 178:500-505.
3. Nakano Y, Muro S, Sakai H, et al. (2000) Computed tomographic measurements of airway dimensions and emphysema in smokers correlation with lung function. American Journal of Respiratory and Critical Care Medicine 162:1102-1108.
4. Coxson HO (2008) Quantitative chest tomography in COPD research: chairman's summary. Proc Am Thorac Soc 5:874-877.
5. Han MK, Kazerooni EA, Lynch DA, et al. (2011) Chronic obstructive pulmonary disease exacerbations in the COPDGene study: associated radiologic phenotypes. Radiology 261:274-282.
6. Muller NL, Staples CA, Miller RR, Abboud RT (1988) "Density mask". An objective method to quantitate emphysema using computed tomography. Chest 94:782-787.
7. Guenard H, Diallo MHH, Laurent F, Vergeret J (1992) Lung density and lung mass in emphysema. Chest 102:198-203.
8. Lamers RJ, Thelissen GR, Kessels AG, Wouters EF, van Engelshoven JM (1994) Chronic obstruc-tive pulmonary disease: evaluation with spirometrically controlled CT lung densitometry. Radiology 193:109-113.
9. Matsuoka S, Yamashiro T, Washko GR, Kurihara Y, Nakajima Y, Hatabu H (2010) Quantitative CT assessment of chronic obstructive pulmonary disease. Radiographics 30:55-66.
10. de Jong PA, Muller NL, Pare PD, Coxson HO (2005) Computed tomographic imaging of the airways: relationship to structure and function. Eur Resp J 26:140-152.
11. Kosciuch J, Krenke R, Gorska K, Zukowska M, Maskey-Warzechowska M, Chazan R (2009) Relationship between airway wall thickness assessed by high-resolution computed tomography and lung function in patients with asthma and chronic obstructive pulmonary disease. Journal of Physiology and Pharmacology 60 Suppl 5:71-76.
12. Washko GR, Criner GJ, Mohsenifar Z, et al. (2008) Computed Tomographic-Based Quantification of Emphysema and Correlation to Pulmonary Function and Mechanics. COPD-Journal of Chronic Obstructive Pulmonary Disease 5:177-186.

13. Akira M, Toyokawa K, Inoue Y, Arai T (2009) Quantitative CT in chronic obstructive pulmonary disease: Inspiratory and expiratory assessment. American Journal of Roentgenology 192:267-272.
14. Matsuoka S, Kurihara Y, Yagihashi K, Nakajima Y (2007) Quantitative assessment of peripheral airway obstruction on paired expiratory/inspiratory thin-section computed tomography in chronic obstructive pulmonary disease with emphysema. Journal of Computer Assisted Tomography 31:384-389.
15. Orlandi I, Moroni C, Camiciottoli G, et al. (2004) Spirometric-gated computed tomography quantitative evaluation of lung emphysema in chronic obstructive pulmonary disease - A comparison of 3 techniques. Journal of Computer Assisted Tomography 28:437-442.
16. Moher D, Liberati A, Tetzlaff J, Altman DG (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. British Medical Journal 339:332-336.
17. Whiting PF, Weswood ME, Rutjes AW, Reitsma JB, Bossuyt PN, Kleijnen J (2006) Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. BMC Medical Research Methodology 6:9.
18. Bon JM, Leader JK, Weissfeld JL, et al. (2009) The influence of radiographic phenotype and smoking status on peripheral blood biomarker patterns in chronic obstructive pulmonary disease. PLoS One 4:e6865.
19. Dransfield MT, Washko GR, Foreman MG, Estepar RSJ, Reilly J, Bailey WC (2007) Gender differences in the severity of CT emphysema in COPD. Chest 132:464-470.
20. Hasegawa M, Nasuhara Y, Onodera Y, et al. (2006) Airflow limitation and airway dimensions in chronic obstructive pulmonary disease. American Journal of Respiratory and Critical Care Medicine 173:1309-1315.
21. Leader JK, Zheng B, Sciurba FC, et al. (2008) The relation of airway size to lung function. Proceedings of SPIE - Medical Imaging 6916:691623.
22. Lee YK, Oh YM, Lee JH, et al. (2008) Quantitative assessment of emphysema, air trapping, and airway thickening on computed tomography. Lung 186:157-165.
23. Park YS, Seo JB, Kim N, et al. (2008) Texture-based quantification of pulmonary emphysema on high-resolution computed tomography: comparison with density-based quantification and correlation with pulmonary function test. Investigative Radiology 43:395-402.
24. Pauls S, Gulkin D, Feuerlein S, et al. (2010) Assessment of COPD severity by computed tomography: correlation with lung functional testing. Clinical Imaging 34:172-178.
25. Washko GR, Dransfield MT, Estepar RS, et al. (2009) Airway wall attenuation: a biomarker of airway disease in subjects with COPD. Journal of Applied Physiology 107:185-191.

26. Yamashiro T, Matsuoka S, Bartholmai BJ, et al. (2010) Collapsibility of lung volume by paired inspiratory and expiratory CT scans: correlations with lung function and mean lung density. Academic Radiology 17:489-495.
27. Zhang J, Yu TL, Wu Q, Ye N, Cao Y, Zhou N (2008) Quantitative MSCT study of lung attenuation in patients with chronic obstructive pulmonary disease. Chinese Journal of Medical Imaging Technology 24:1611-1614.
28. Yamashiro T, Matsuoka S, San Jose Estepar R, et al. (2010) Quantitative assessment of bronchial wall attenuation with thin-section CT: an indicator of airflow limitation in chronic obstructive pulmonary disease. American Journal of Roentgenology 195:363-369.
29. Hesselbacher SE, Ross R, Schabath MB, et al. (2011) Cross-sectional analysis of the utility of pulmonary function tests in predicting emphysema in ever-smokers. International Journal of Environmental Research and Public Health 8:1324-1340.
30. Ohno Y, Koyama H, Yoshikawa T, et al. (2011) Comparison of capability of dynamic O2-enhanced MRI and quantitative thin-section MDCT to assess COPD in smokers. European Journal of Radiology DOI:10.1016/j.ejrad.2011.1002.1004.
31. Iwasawa T, Takahashi H, Ogura T, et al. (2011) Influence of the distribution of emphysema on diaphragmatic motion in patients with chronic obstructive pulmonary disease. Japanese Journal of Radiology 29:256-264.
32. Jin LF, Miao JT, Li ZY, Xie XQ, Dai XM (2007) Study of voxel index from multislice CT in evaluating chronic obstructive pulmonary disease air-trapping. Chinese Journal of Medical Imaging Technology 23:1033-1036.
33. Falaschi F, Miniati M, Battolla L, et al. (1995) Quantification of pulmonary emphysema with computerized tomography: comparison with various methods. La Radiologia medica 90:16-23.
34. Deveci F, Murat A, Turgut T, Altuntas E, Muz MH (2004) Airway wall thickness in patients with COPD and healthy current smokers and healthy non-smokers: assessment with high resolution computed tomographic scanning. Respiration 71:602-610.
35. Iwasawa T, Takahashi H, Ogura T, et al. (2007) Correlation of lung parenchymal MR signal intensity with pulmonary function tests and quantitative computed tomography (CT) evaluation: a pilot study. Journal of Magnetic Resonance Imaging 26:1530-1536.
36. Heussel CP, Herth FJ, Kappes J, et al. (2009) Fully automatic quantitative assessment of emphysema in computed tomography: comparison with pulmonary function testing and normal values. European Radiology 19:2391-2402.
37. Shaker SB, Maltbaek N, Brand P, Haeussermann S, Dirksen A (2005) Quantitative computed tomography and aerosol morphometry in COPD and alpha-antitrypsin deficiency. European Respiratory Journal 25:23-30.

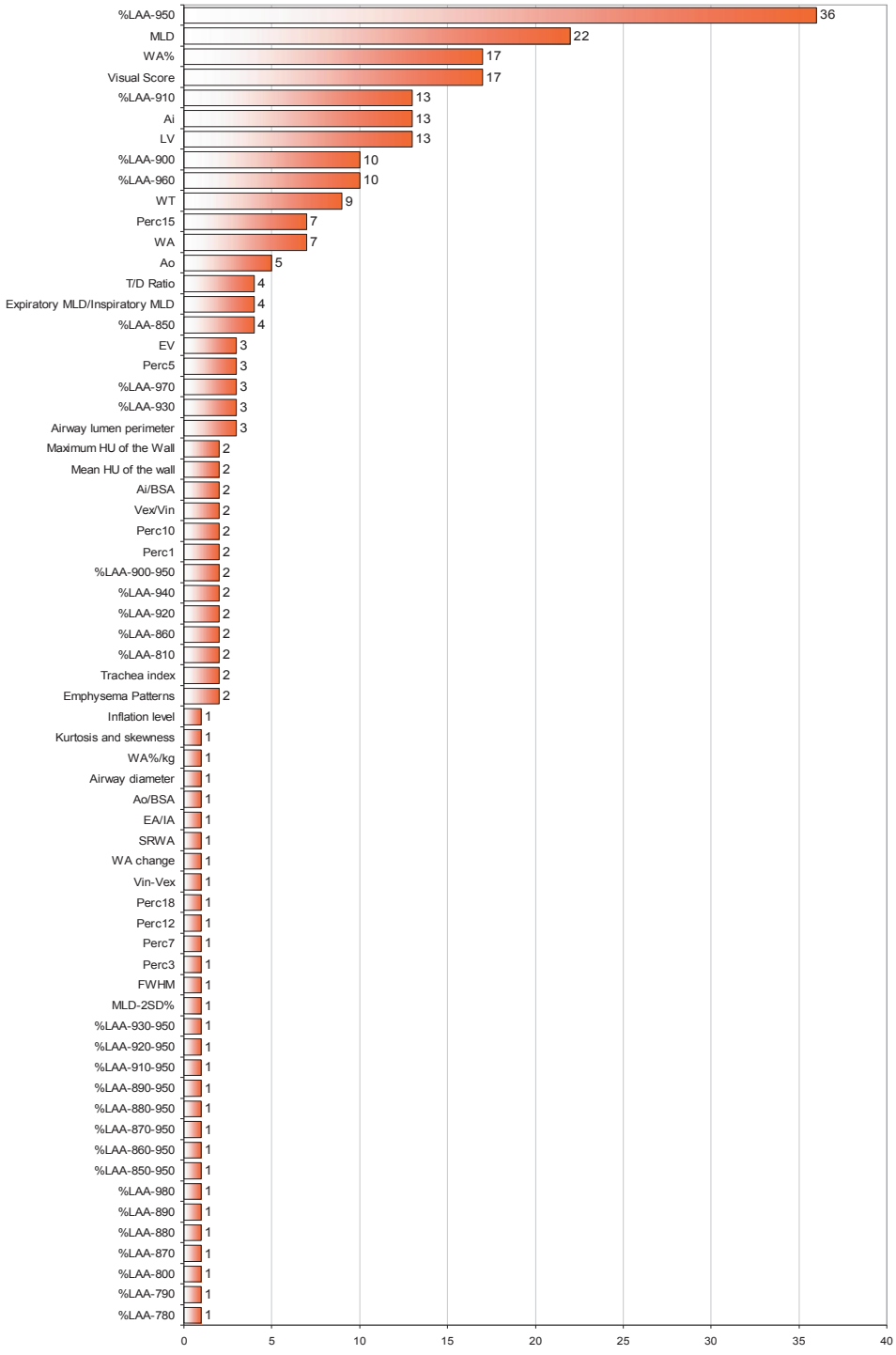
38. Hogg JC, Chu F, Utokaparch S, et al. (2004) The nature of small-airway obstruction in chronic obstructive pulmonary disease. N Engl J Med 350:2645-2653.
39. Camiciottoli G, Bartolucci M, Maluccio NM, et al. (2006) Spirometrically gated high-resolution CT findings in COPD - Lung attenuation vs lung function and dyspnea severity. Chest 129:558-564.
40. O'Donnell RA, Peebles C, Ward JA, et al. (2004) Relationship between peripheral airway dysfunction, airway obstruction, and neutrophilic inflammation in COPD. Thorax 59:837-842.
41. Zaporozhan J, Ley S, Eberhardt R, et al. (2005) Paired inspiratory/expiratory volumetric thin-slice CT scan for emphysema analysis: comparison of different quantitative evaluations and pulmo-nary function test. Chest 128:3212-3220.
42. Achenbach T, Weinheimer O, Biedermann A, et al. (2008) MDCT assessment of airway wall thickness in COPD patients using a new method: Correlations with pulmonary function tests. European Radiology 18:2731-2738.
43. Boschetto P, Quintavalle S, Zeni E, et al. (2006) Association between markers of emphysema and more severe chronic obstructive pulmonary disease. Thorax 61:1037-1042.
44. Newell JD, Jr. (2008) Quantitative computed tomography of lung parenchyma in chronic obstruc-tive pulmonary disease: an overview. Proceedings of the American Thoracic Society 5:915-918.
45. Mohamed Hoesein FAA, De Hoop B, Zanen P, et al. (2011) CT-quantified emphysema in male heavy smokers: association with lung function decline. Thorax 66:782-787.
46. Bae KT, Slone RM, Gierada DS, Yusen RD, Cooper JD (1997) Patients with emphysema: quantitative CT analysis before and after lung volume reduction surgery. Radiology 203:705-714.
47. Bafadhel M, Umar I, Gupta S, et al. (2011) The role of computed tomography in multi-dimensional phenotyping of chronic obstructive pulmonary disease. Chest 140:634-642.
48. Baldi S, Miniati M, Bellina CR, et al. (2001) Relationship between extent of pulmonary emphysema by high-resolution computed tomography and lung elastic recoil in patients with chronic obstructive pulmonary disease. American Journal of Respiratory and Critical Care Medicine 164:585-589.
49. Beinert T, Behr J, Mehnert F, et al. (1995) Spirometrically controlled quantitative CT for assessing diffuse parenchymal lung disease. Journal of Computer Assisted Tomography 19:924-931.
50. Cavigli E, Camiciottoli G, Diciotti S, et al. (2009) Whole-lung densitometry versus visual assessment of emphysema. European Radiology 19:1686-1692.
51. Cerveri I, Dore R, Corsico A, et al. (2004) Assessment of emphysema in COPD - a functional and radiologic study. Chest 125:1714-1718.

52. Crausman RS, Ferguson G, Irvin CG, Make B, Newell JD (1995) Quantitative chest computed tomography as a means of predicting exercise performance in severe emphysema. Academic Radiology 2:463-469.
53. Daghfous J, Beji M, Kanoun N, et al. (1993) Pulmonary emphysema quantification using computed tomography and correlations with respiratory function tests. Revue des Maladies Respiratoires 10:299-305.
54. D'Anna SE, Asnaghi R, Caramori G, et al. (2011) High-resolution computed tomography quantitation of emphysema is correlated with selected lung function values in stable COPD. Respiration DOI:10.1159/000329871.
55. Demir T, Ikitimur H, Akpinar Tekgunduz S, et al. (2005) The relationship between pulmonary function tests, thorax HRCT, and quantitative ventilation-perfusion scintigraphy in chronic obstructive pulmonary disease. Tüberkülozve Toraks Dergisi 53:347-353.
56. Gelb AF, Schein M, Kuei J, et al. (1993) Limited contribution of emphysema in advanced chronic obstructive pulmonary disease. American Review of Respiratory 147:1157-1161.
57. Grydeland TB, Dirksen A, Coxson HO, et al. (2010) Quantitative computed tomography measures of emphysema and airway wall thickness are related to respiratory symptoms. American Journal of Respiratory and Critical Care Medicine 181:353-359.
58. Grydeland TB, Thorsen E, Dirksen A, et al. (2011) Quantitative CT measures of emphysema and airway wall thickness are related to D(L)CO. Respir Med 105:343-351.
59. Jogi J, Ekberg M, Jonson B, Bozovic G, Bajc M (2011) Ventilation/perfusion SPECT in chronic obstructive pulmonary disease: an evaluation by reference to symptoms, spirometric lung function and emphysema, as assessed with HRCT. Eur J Nucl Med Mol Imaging 38:1344-1352.
60. Kim WJ, Silverman EK, Hoffman E, et al. (2009) CT metrics of airway disease and emphysema in severe COPD. Chest 136:396-404.
61. Leader JK, Bin Z, Fuhrman CR, et al. (2009) Association between lung function and airway wall density. Proceedings of the SPIE-The International Society for Optical Engineering 72622J (72629 pp.).
62. Lee JS, Ra SW, Chae EJ, et al. (2011) Validation of the Lower Limit of Normal Diffusing Capacity for Detecting Emphysema. Respiration 81:287-293.
63. Lee HJ, Seo JB, Chae EJ, et al. (2011) Tracheal morphology and collapse in COPD: Correlation with CT indices and pulmonary function test. European Journal of Radiology 80:e531-e535.

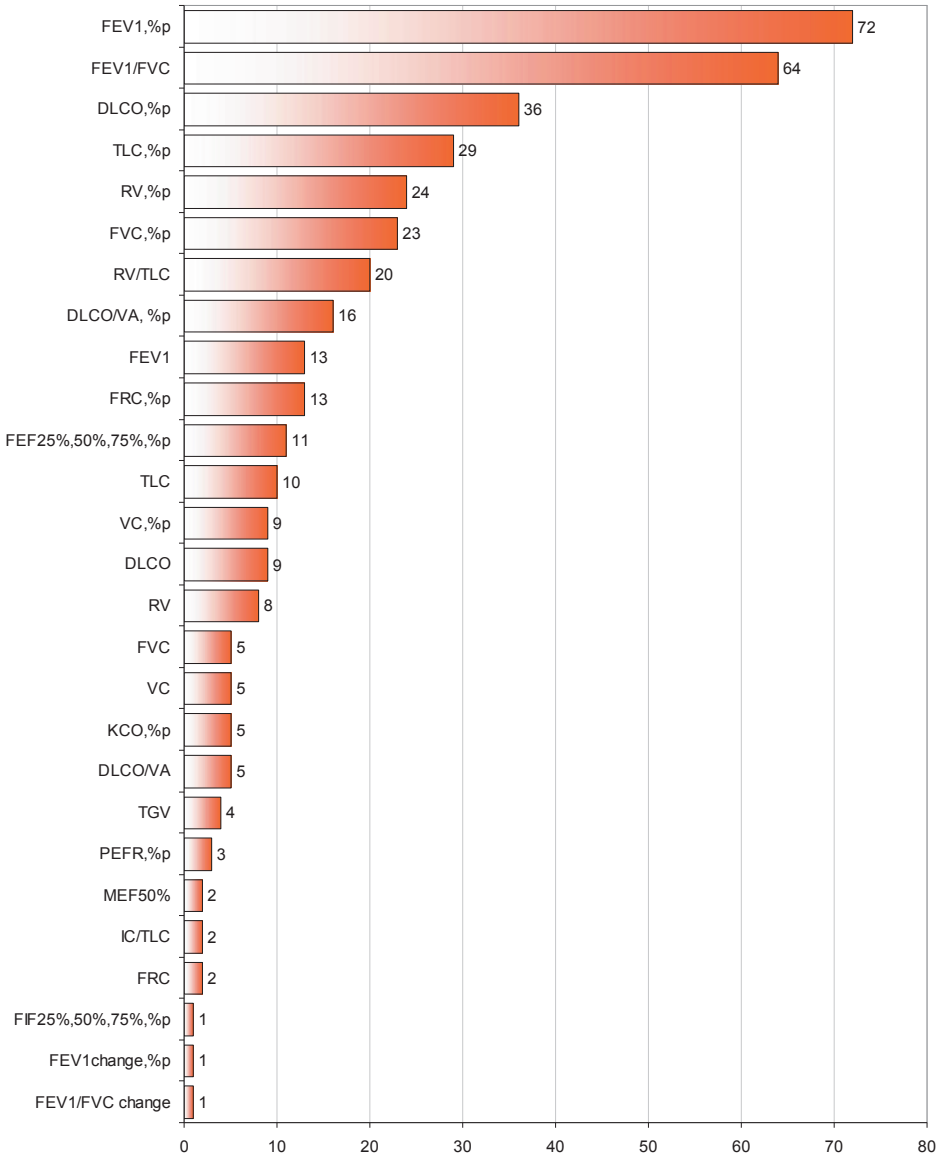
64. Li S, Chen L, Du JM, Hua SC (2009) Application of measurement of lung volumes from MSCT images in evaluation of pulmonary function of patients with chronic obstructive pulmonary disease. Journal of Jilin University Medicine Edition 35:1167-1171.
65. Madani A, Van Muylem A, Gevenois PA (2010) Pulmonary emphysema: effect of lung volume on objective quantification at thin-section CT. Radiology 257:260-268.
66. Marquez-Martin E, Ramos PC, Lopez-Campos JL, et al. (2011) Components of physical capacity in patients with chronic obstructive pulmonary disease: relationship with phenotypic expression. Int J Chron Obstruct Pulmon Dis 6:105-112.
67. Matsuoka S, Kurihara Y, Yagihashi K, Hoshino M, Nakajima Y (2008) Airway dimensions at inspiratory and expiratory multisection CT in chronic obstructive pulmonary disease: correlation with airflow limitation. Radiology 248:1042-1049.
68. Matsuoka S, Kurihara Y, Yagihashi K, Hoshino M, Watanabe N, Nakajima Y (2008) Quantitative assessment of air trapping in chronic obstructive pulmonary disease using inspiratory and expiratory volumetric MDCT. American Journal of Roentgenology 190:762-769.
69. Mets OM, Murphy K, Zanen P, et al. (2011) The relationship between lung function impairment and quantitative computed tomography in chronic obstructive pulmonary disease. European Radiology DOI:10.1007/s00330-00011-02237-00339.
70. Mishima M, Itoh H, Sakai H, et al. (1999) Optimized scanning conditions of high resolution CT in the follow-up of pulmonary emphysema. Journal of Computer Assisted Tomography 23:380-384
71. Moron K, Liebhart J, Liebhart E, Ciesielska A, Badowski R (2004) Structural and functional lung changes in patients with COPD and asthma - The benefit of quantitative HRCT assessment. Advances in Clinical and Experimental Medicine 13:59-66.
72. Moroni C, Mascalchi M, Bartolucci M, Camiciottoli G, Pistolesi M, Villari N (2001) High resolution and spirometric synchronization computerized tomography in chronic obstructive bronchopneumopathy. La Radiologia medica 101:25-30.
73. Nakano Y, Sakai H, Muro S, et al. (1999) Comparison of low attenuation areas on computed tomographic scans between inner and outer segments of the lung in patients with chronic obstructive pulmonary disease: incidence and contribution to lung function. Thorax 54:384-389.
74. Ohara T, Hirai T, Sato S, et al. (2006) Comparison of airway dimensions in different anatomic locations on chest CT in patients with COPD. Respirology 11:579-585.
75. Orlandi I, Moroni C, Camiciottoli G, et al. (2005) Chronic obstructive pulmonary disease: thin-section CT measurement of airway wall thickness and lung attenuation. Radiology 234:604-610.

76. Pescarolo M, Sverzellati N, Verduri A, et al. (2008) How much do GOLD stages reflect CT abnormalities in COPD patients? Radiologia Medica 113:817-829.
77. Petersen J, Pechin L, Nielsen M, et al. (2010) Quantitative analysis of airway abnormalities in CT. Proceedings of the SPIE - The International Society for Optical Engineering 76241S (76212 pp.).
78. Sandek K, Bratel T, Lagerstrand L, Rosell H (2002) Relationship between lung function, ventilation-perfusion inequality and extent of emphysema as assessed by high-resolution computed tomography. Respiratory Medicine 96:934-943.
79. Scichilone N, La Sala A, Bellia M, et al. (2008) The airway response to deep inspirations decreases with COPD severity and is associated with airway distensibility assessed by computed tomography. Journal of Applied Physiology 105:832-838.
80. Sorensen L, Shaker SB, de Bruijne M (2010) Quantitative analysis of pulmonary emphysema using local binary patterns. IEEE Transactions on Medical Imaging 29:559-569.
81. Spiropoulos K, Trakada G, Kalamboka D, et al. (2003) Can high resolution computed tomography predict lung function in patients with chronic obstructive pulmonary disease? Lung 181:169-181.
82. de Torres JP, Bastarrika G, Zagaceta J, et al. (2011) Emphysema presence, severity, and distribution has little impact on the clinical presentation of a cohort of patients with mild to moderate COPD. Chest 139:36-42.
83. Tsushima K, Sone S, Fujimoto K, et al. (2010) Identification of occult parechymal disease such as emphysema or airway disease using screening computed tomography. COPD-Journal of Chronic Obstructive Pulmonary Disease 7:117-125.
84. van der Lee I, van Es HW, Noordmans HJ, van den Bosch JM, Zanen P (2006) Alveolar volume determined by single-breath helium dilution correlates with the high-resolution computed tomography-derived nonemphysematous lung volume. Respiration 73:468-473.
85. Watanuki Y, Suzuki S, Nishikawa M, et al. (1994) Correlation of quantitative CT with selective alveolobronchogram and pulmonary function tests in emphysema. Chest 106:806-813.
86. Yamashiro T, Matsuoka S, Estepar RSJ, et al. (2011) Kurtosis and skewness of density histograms on inspiratory and expiratory CT scans in smokers. COPD: Journal of Chronic Obstructive Pulmonary Disease 8:13-20.
87. Zompatori M, Battaglia M, Rimondi MR, et al. (1997) Quantitative assessment of pulmonary emphysema with computerized tomography: comparison of the visual score and high resolution computerized tomography, expiratory density mask with spiral computerized tomography and respiratory function tests. La Radiologia medica 93:374-381.

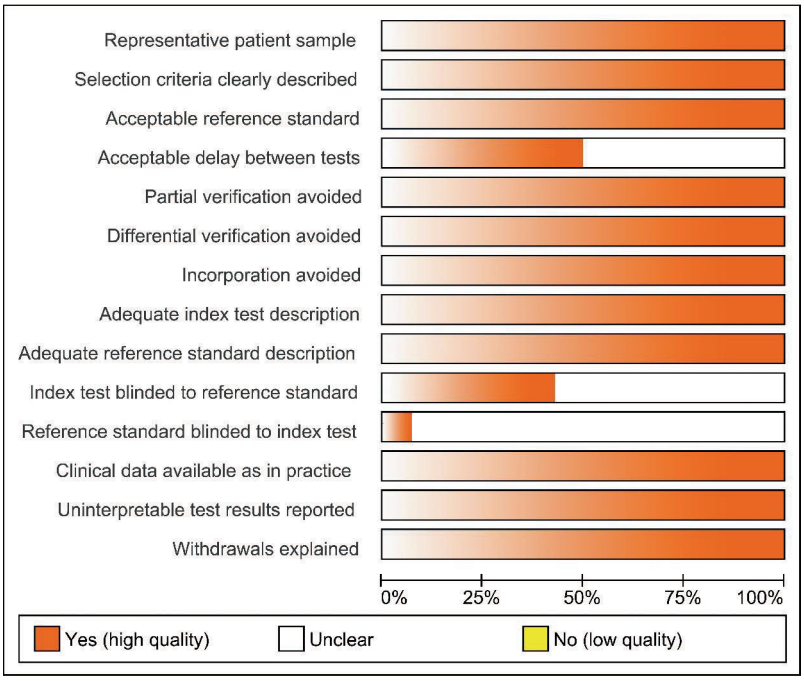
88. Zompatori M, Fasano L, Battista G, Pacilli AM (2001) Longitudinal follow-up evaluation of pulmonary emphysema using high-resolution computed tomography (HRCT) study and pulmo-nary function tests. La Radiologia medica 102:122-126.
89. Zompatori M, Fasano L, Pacilli AM, Battista G, Cavina M, Pezzi S (2001) Automatic evaluation of total lung capacity and of emphysema involvement with spiral computerized tomography (CT) in obstructive pneumonia. La Radiologia medica 101:18-24.
90. Zompatori M, Fasano L, Mazzoli M, et al. (2002) Spiral CT evaluation of pulmonary emphysema using a low-dose technique. La Radiologia medica 104:13-24.


%LAA-950 36
MLD 22
WA% 17
Visual Score 17
%LAA-910 13
Ai 13
LV 13
%LAA-900 10
%LAA-960 10
WT 9
Perc15 7
WA 7
Ao 5
T/D Ratio 4
Expiratory MLD/Inspiratory MLD 4
%LAA-850 4
EV 3
Perc5 3
%LAA-970 3
%LAA-930 3
Airway lumen perimeter 3
Maximum HU of the Wall 2
Mean HU of the wall 2
Ai/BSA 2
Vex/Vin 2
Perc10 2
Perc1 2
%LAA-900-950 2
%LAA-940 2
%LAA-920 2
%LAA-860 2
%LAA-810 2
Trachea index 2
Emphysema Patterns 2
Inflation level 1
Kurtosis and skewness 1
WA%/kg 1
Airway diameter 1
Ao/BSA 1
EA/IA 1
SRWA 1
WA change 1
Vin-Vex 1
Perc18 1
Perc12 1
Perc7 1
Perc3 1
FWHM 1
MLD-2SD% 1
%LAA-930-950 1
%LAA-920-950 1
%LAA-910-950 1
%LAA-890-950 1
%LAA-880-950 1
%LAA-870-950 1
%LAA-860-950 1
%LAA-850-950 1
%LAA-980 1
%LAA-890 1
%LAA-880 1
%LAA-870 1
%LAA-800 1
%LAA-790 1
%LAA-780 1
0 5 10 15 20 25 30 35 40

◄ **Supplemental Figure A** Counting of CT measurements in the systematic review. %LAA = percentage low attenuation area; MLD = mean lung density; WA = airway wall area; Ai = airway lumen area; LV = lung volume; WT = wall thickness; Perc = percentile point of lung density; Ao = total airway area; T/D radio = radio of airway wall thickness to total diameter; EV = emphysema volume; HU = Hounsfield unit; BSA = body surface area; Vex = expiration volume; Vin = inspiration volume; EA = expiration airway lumen area; IA = inspiration airway lumen area; SRWA = square root of wall area; FWHM = full width at half maximum; SD = standard deviation.



Supplemental Figure B Counting of pulmonary function test parameters in the systematic review. %pred = predicted percentage; FEV_1 = forced expiratory volume in the first second; FEF = forced expiratory flow; VA = alveolar volume; TLC = total lung capacity; RV = residual volume; DLCO = diffusing capacity of the lung for carbon monoxide; VC = vital capacity; TGV = thoracic gas volume; FVC = forced vital capacity; PEFR = peak expiratory flow rate; KCO = carbon monoxide transfer coefficient; MEF = maximal expiratory flow; IC = inspiratory capacity; FRC = functional residual capacity; FIF = forced inspiratory flow.



Supplemental Figure C Study quality summaries of articles included in the meta-analysis, assessed by the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool.

Supplemental Table A Literature search strategy. Last search on January 5, 2012

Search terms used to identify relevant citations

A: Computed tomography
CT
B: Lung function*
Respiratory function*
Pulmonary function*
Spirometr*
Diffusing capacity
Diffusion capacity
Airway obstruction parameter*
C: Chronic obstructive pulmonary disease*
COPD
Chronic obstructive lung disease*
Chronic airflow obstruction*
To search: A and B and C
The term 'CT' is restrained to the title and abstract
Publication date: From January 1976 to December 2011

Search query for PubMed
#1: ("Tomography, X-Ray Computed"[MeSH] OR "computed tomography"[TIAB] OR CT[TIAB])
#2: ("Pulmonary Disease, Chronic Obstructive"[MeSH] OR COPD OR "chronic obstructive pulmonary disease*" OR "chronic obstructive lung disease*" OR "Chronic airflow obstruction*")
#3: ("Pulmonary Function Test" OR Spirometr* OR "pulmonary function*" OR "Lung function*" OR "Respiratory function*" OR "Diffusing capacity" OR "Diffusion capacity" OR "Airway obstruction parameter*")
Grammar in advanced search: 1976/01:2011/12 [dp] and #1 and #2 and #3

Search query for Embase
#1: ((Computed tomography):ab,ti OR CT:ab,ti)
#2: ((Chronic Obstructive Pulmonary Disease*) OR COPD OR (Chronic Obstructive Lung Disease*) OR (Chronic obstructive airway disease*) OR (Chronic airflow obstruction*))
#3: ('lung function test'/exp OR (Spirometr*) OR ('Lung function') OR ('Pulmonary function') OR ('Respiratory function') OR ('Lung functions') OR ('Pulmonary functions') OR ('Respiratory functions') OR ('Diffusing capacity') OR ('Diffusion capacity') OR ('Airway obstruction'))
Grammar in advanced search: #1 AND #2 AND #3 AND [1-1-1976]/sd NOT
(#1 AND #2 AND #3 AND [31-12-2011]/sd)

Search query for Web of Knowledge
#1 topic: ((Computed tomography) OR CT)
#2 topic: ((Chronic Obstructive Pulmonary Disease*) OR COPD OR (Chronic Obstructive Lung Disease*) OR (Chronic obstructive airway disease*) OR (Chronic airflow obstruction*))
#3 topic: (Spirometr* OR (Lung function*) OR (Pulmonary function*) OR (Respiratory function*) OR (Diffusing capacity) OR (Diffusion capacity) OR (Airway obstruction parameter*))
Grammar: #1 topic and #2 topic and #3 topic

Supplemental Table B Characteristics of studies included in the systematic review

Study, Year	Patients, *n*	Men, %	Age, Year ± SD or year (range)	COPD Severity: GOLD stage or FEV_1%pred±SD or FEV_1%pred (range)	CT type, slice	Inspiratory or expiratory CT	Volumetric CT examination	Radiation Dose
Achenbach 2008 [42]	16	75	65(50-75)	Stage II-III	4	Insp	Vol	Nor
Akira 2009 [13]*	76	88	67(37-85)	Stage 0-IV	16	Both	Vol	Nor
Bae 1997 [46]	10	NA	57(41-77)	NA	1	Both	N-vol	Nor
Bafadhel 2011 [47]	75	77	67(43-88)	47±2%†	16	Insp	N-vol	Nor
Baldi 2001 [48]	24	75	61±11	35%(17-72%)	1	Insp	N-vol	Nor
Beinert 1995 [49]	11	82	48(34-56)	41±0.04%†	1	Both	N-vol	Nor
Bon 2009 [18]*	234	50	61(50-78)	Stage 0-IV	4&8	Insp	Vol	Low
Boschetto 2006 [43]	26	85	71±3(*n*=11)† 70±2(*n*=15)†	31±2.6%(*n*=11)† 47±3.8%(*n*=15)†	1	Insp	N-vol	Nor
Camiciottoli 2006[39]	51	90	64(43-78)	52%(15-106%)	1	Both	N-vol	Nor
Cavigli 2009 [50]	30	77	68(52-81)	Stage I-IV	16	Insp	Vol	Nor
Cerveri 2004 [51]	39	90	64(48-80)	47%(21-72%)	1	Exp	N-vol	Nor
Crausman 1995 [52]	9	22	67±2†	40%(20-89%)	1	Insp	N-vol	Nor
Daghfous 1993 [53]	51	90	55(25-80)	40±21%(*n*=31) 39±17%(*n*=20)	1	Insp	N-vol	NA
D'Anna 2011 [54]	59	68	68±7	52±18	16	Insp	N-vol	Nor
Demir 2005 [55]	16	100	65±6	41±15%	1	Insp	N-vol	NA
Deveci 2004 [34]	22	100	57±5(*n*=17) 61±7(*n*=5)	Stage II-III	1	Insp	N-vol	Nor
Dransfield 2007 [19]*	396	62	63±5(*n*=246) 61±5(*n*=150)	Stage 0-IV	Multi	Insp	Vol	Low
Falaschi 1995 [33]	46	83	63(46-78)	NA	1	Both	N-vol	Nor
Gelb 1993 [56]	56	48	68(53-76)	46%(18-79%)	1	Insp	N-vol	Nor
Grydeland 2010 [57]	463	65	65±9(*n*=299) 63±9(*n*=164)	Stage II-IV	8	Insp	N-vol	Nor
Grydeland 2011 [58]	288	70	64±10(*n*=202) 60±8(*n*=86)	Stage II-IV	8	Insp	N-vol	Nor
Hasegawa 2006 [20]*	52	96	72(41-84)	Stage I-IV	4	Insp	Vol	Nor
Hesselbacher2011[29]*	224	65	>40	Stage I-IV	64	Insp	Vol	Nor
Heussel 2009 [36]	102	NA	64(20-87)	Stage III-IV	4	Insp	Vol	Nor
Iwasawa 2007 [35]	19	100	71±7	Stage II-IV	16	Insp	Vol	Nor
Iwasawa 2011 [31]*	35	100	70±6	Stage I-IV	16	Insp	Vol	Nor
Jin 2007 [32]	43	51	65(45-85)	NA	16	Both	Vol	Nor
Jogi 2011 [59]	30	65	65(53-76)	51%(25-81%)	Multi	Insp	Vol	NA
Kim 2009 [60]	338	64	68±6	<45%	1	Insp	N-vol	Nor
Kosciuch 2009 [11]	12	58	57±9	72±19%	16	Insp	N-vol	Nor
Lamers 1994 [8]	40	83	60±8(*n*=20) 70±7(*n*=20)	46±12%(*n*=20) 51±14%(*n*=20)	1	Both	N-vol	Nor
Leader 2008 [21]*	240	NA	NA	Stage 0-IV	4(*n*=112) 8(*n*=128)	Insp	Vol	Low
Leader 2009 [61]	200	NA	NA	Stage 0-IV	64	Insp	Vol	Nor
Lee 2008 [22]*	34	97	65(50-78)	45%(17-82%)	16	Both	Vol	Nor
Lee 2011 [62]	197	96	67±7(*n*=126) 65±8(*n*=71)	44±15%(*n*=126) 55±15%(*n*=71)	16	Both	Vol	Nor
Lee 2011 [63]	115	NA	65	Stage I-IV	16	Both	Vol	Nor
Li 2009 [64]	24	75	56(32-70)	56±16%	64	Both	N-vol	Nor
Madani 2010 [65]	16	63	62(48-83)	Stage I-IV	4	Both	N-vol	Nor
Marquez-Martin2011[66]	64	NA	64±7	Stage I-IV	NA	Insp	N-vol	Nor

Matsuoda 2007 [14]	32	88	73(57-89)	Stage 0-IV	1	Both	N-vol	Nor
Matsuoda 2008 [67]	50	80	70(57-89)	Stage I-IV	64	Both	Vol	Nor
Matsuoda 2008 [68]	36	86	71(57-89)	Stage I-IV	64	Both	Vol	Nor
Mets 2011 [69]	198	NA	about 60	Stage I-IV	16	Both	Vol	Low
Mishima 1999 [70]	72	NA	NA	NA	1	Insp	N-vol	Both
Mohamed Hoesein 2011 [45]	2085	2085	60±5	Stage 0-III	16	Insp	Vol	Low
Moron 2004 [71]	16	63	62±9	40±18%	1	Insp	N-vol	NA
Moroni 2001 [72]	20	95	63(42-73)	NA	1	Both	N-vol	Nor
Nakano 1999 [73]	73	NA	69±6	46±20%	1	Insp	N-vol	Nor
Nakano 2000 [3]	94	NA	NA	48%(8-124%)	1	Insp	N-vol	Nor
O'Donnel 2004 [40]	44	NA	50±7(*n*=17) 57±7(*n*=10) 55±7(*n*=17)	Stage 0-IV	1	Both	N-vol	Nor
Ohara 2006 [74]	30	100	69±8	41±16%	1	Insp	N-vol	Nor
Ohno 2011 [30]*	186	65	(23-87)	Stage 0-IV	16&64	Insp	Vol	Nor
Orlandi 2004 [15]	11	82	68(60-75)	NA	1	Insp	N-vol	Both
Orlandi 2005 [75]	42	88	63(42-73)	49%(15-83%)	1	Insp	N-vol	Nor
Park 2008 [23]*	39	92	66(51-79)	44±15%	16	Insp	Vol	Nor
Patel 2008 [2]	519	51	58±5	<60%	1	Insp	N-vol	Nor
Pauls 2010 [24]*	145	NA	NA	Stage I-IV	16	Insp	Vol	Nor
Pescarolo 2008 [76]	43	58	62(44-81)	Stage 0-IV	16&64	Insp	Vol	Nor
Petersen 2010 [77]	152	NA	NA	≥Stage II	16	Insp	Vol	Low
Sandek 2002 [78]	20	40	60±8	38±16%	1	Both	N-vol	Nor
Scichilone 2008 [79]	15	NA	69(53-90)	Stage I-IV	40	Insp	Vol	Nor
Shaker 2005 [37]	42	38	63±8	48±13%	4	Insp	Vol	Low
Sorensen 2010 [80]	20	NA	64(49-80)	57%(37-76%)	4	Insp	Vol	Nor
Spiropoulos 2003 [81]	20	90	59±9	57±26%	1	Both	N-vol	NA
Torres 2011 [82]	115	84	63±10	75±15%	64	Insp	Vol	Low
Tsushima 2010 [83]	48	83	61±9	Stage I-III	4	Insp	N-vol	Low
Van Der Lee 2006[84]	50	58	60(29-83)	46±24%	1	Insp	N-vol	Nor
Washko 2008 [12]	1094	61	67±6	<45%	1	Insp	N-vol	Nor
Washko 2009 [25]*	224	42	62±5	Stage I-IV	4	Insp	Vol	Low
Watanuki 1994 [85]	21	NA	65(38-77)	<70%	1	Insp	N-vol	NA
Yamashiro 2010 [26]*	46	57	68(46-81)	Stage 0-IV	16	Both	Vol	Nor
Yamashiro 2010 [28]*	114	57	62(56-74)	Stage I-IV	4	Insp	Vol	Low
Yamashiro 2011 [86]	46	57	68(46-81)	Stage 0-IV	16	Both	Vol	Nor
Zampatori 1997 [87]	20	80	69(61-86)	33%(21-57%)	1	Both	N-vol	Nor
Zampatori 2001 [88]	15	80	63	32%(22-63%)	1	Insp	N-vol	Nor
Zampatori 2001 [89]	17	77	66(47-78)	47±25%	1	Insp	N-vol	Nor
Zampatori 2002 [90]	18	67	61(27-81)	<71%	4	Insp	N-vol	Nor
Zaporozhan 2005 [41]	31	71	60(41-76)	Stage II-IV	16	Both	Vol	Nor
Zhang 2008 [27]*	50	66	67±10	Stage 0-IV	16	Insp	Vol	Nor

SD = standard deviation; SE = standard error; Insp = inspiratory; exp = expiratory; Vol = volumetric scan; N-vol = non-volumetric scan; Nor = normal dose; GOLD = the Global Initiative for Chronic Obstructive Lung Disease; FEV_1 %pred = predicted forced expiratory volume in the first second; PFT = pulmonary function test; NA = not available; VC = vital capacity.

Supplemental Table C Correlation coefficients between CT measurements and airflow obstruction parameters of pulmonary function test in the systematic review

Study, Year	CT Measurements	Correlation coefficients between CT measurements and airflow obstruction, *r*	
		Inspiratory scan	Expiratory scan
Achenbach 2008 [42]	WA%, WA	WA% & FEV_1 %pred : -0.537 WA & FEV_1 %pred : -0.423	NA
Akira 2009 [13]*	%LAA-950, -910, MLD, Perc15 Visual score, etc.	%LAA-950 & FEV_1 %pred: -0.659 %LAA-950 & FEV_1/FVC: -0.712 %LAA-910 & FEV_1 %pred: -0.601 %LAA-910 & FEV_1/FVC: -0.661 MLD & FEV_1 %pred: 0.694 MLD & FEV_1/FVC: 0.764 Perc15 & FEV_1 %pred: 0.292 Perc15 & FEV_1/FVC: 0.600	%LAA-950 & FEV_1 %pred: -0.668 %LAA-950 & FEV_1/FVC: -0.666 %LAA-910 & FEV_1 %pred: -0.632 %LAA-910 & FEV_1/FVC: -0.642 MLD & FEV_1 %pred: 0.792 MLD & FEV_1/FVC: 0.721 Perc15 & FEV_1 %pred: 0.352 Perc15 & FEV_1/FVC: 0.544
Bae 1997 [46]	%LAA-900	%LAA-900 & FEV_1 %pred: -0.52 %LAA-900 & FEV_1/FVC: -0.33	%LAA-900 & FEV_1 %pred: -0.87 %LAA-900 & FEV_1/FVC: -0.45
Bafadhel 2011 [47]	Emphysema pattern, WT	No CT quantitative measurements	NA
Baldi 2001 [48]	%LAA-950, MLD	%LAA-950 & FEV_1 %pred: -0.50 MLD & FEV_1 %pred: 0.62	NA
Beinert 1995 [49]	MLD	No reports on the correlation between MLD & FEV_1 %pred or FEV_1/FVC	NA
Bon 2009 [18]*	%LAA-950, WA%	%LAA-950 & FEV_1 %pred: -0.43 WA% & FEV_1 %pred : -0.39	NA
Boschetto 2006 [43]	%LAA-950, MLD	%LAA-950 & FEV_1 %pred: -0.480 MLD & FEV_1 %pred: 0.659	NA
Camiciottoli 2006 [39]	%LAA-950, %LAA-910, MLD	%LAA-910 & FEV_1/FVC: -0.57 MLD & FEV_1/FVC: 0.60	%LAA-910 & FEV_1/FVC: -0.72 MLD & FEV_1/FVC: 0.80
Cavigli 2009 [50]	%LAA-950, MLD, Perc15, etc.	No reports on the correlation to FEV_1 %pred or FEV_1/FVC	NA
Cerveri 2004 [51]	%LAA-900	NA	%LAA-900 & FEV_1 %pred: -0.52
Crausman 1995 [52]	%LAA-900	%LAA-900 & FEV_1 %pred: 0.34† %LAA-900 & FEV_1/FVC: 0.46†	NA
Daghfous 1993 [53]	Visual Score	No CT quantitative measurements	NA
D'Anna 2011 [54]	Visual Score	No CT quantitative measurements	NA
Demir 2005 [55]	Visual Score	No CT quantitative measurements	NA
Deveci 2004 [34]	WA%, T/D ratio	WA% & FEV_1 %pred: -0.713 WA% & FEV_1/FVC: -0.573 T/D ratio & FEV_1 %pred: -0.735 T/D ratio & FEV_1/FVC: -0.579	NA
Dransfield 2007 [19]*	%LAA-950	%LAA-950 & FEV_1 %pred: -0.44 (total) %LAA-950 & FEV_1/FVC: -0.58 (total) %LAA-950 & FEV_1 %pred: -0.42 (men) %LAA-950 & FEV_1/FVC: -0.62 (men) %LAA-950 & FEV_1 %pred:-0.49(women) %LAA-950 & FEV_1/FVC: -0.55 (women)	NA
Falaschi 1995 [33]	%LAA-900, MLD, Visual Score	%LAA-900 & FEV_1 %pred: -0.80 %LAA-900 & FEV_1/FVC: -0.86 MLD & FEV_1 %pred: 0.76 MLD & FEV_1/FVC: 0.81	%LAA-900 & FEV_1 %pred: -0.82 %LAA-900 & FEV_1/FVC: -0.86 MLD & FEV_1 %pred: 0.85 MLD & FEV_1/FVC: 0.89
Gelb 1993 [56]	Visual Score	No CT quantitative measurements	NA
Grydeland 2010 [57]	%LAA-950, -910, etc.	No reports on the correlation to FEV_1 %pred or FEV_1/FVC	NA
Grydeland 2011 [58]	%LAA-950, WT	No reports on the correlation to FEV_1 %pred or FEV_1/FVC	NA
Hasegawa 2006 [20]*	WA%, Ai, WA	WA% & FEV_1 %pred: -0.547 Ai & FEV_1 %pred: 0.731	NA
Hesselbacher 2011 [29]*	%LAA-950, etc.	%LAA-950 & FEV_1/FVC: -0.71 (current smoker) %LAA-950 & FEV_1/FVC: -0.78 (former smoker)	NA
Heussel 2009 [36]	%LAA-950, MLD, Perc15, LV, etc.	%LAA-950 & FEV_1 %pred : -0.35 %LAA-950 & FEV_1/FVC: -0.63 MLD & FEV_1 %pred : 0.43 MLD & FEV_1/FVC: 0.69 Perc15 & FEV_1 %pred : 0.34 Perc15 & FEV_1/FVC: 0.62 LV & FEV_1 %pred : -0.02 LV & FEV_1/FVC: -0.58	NA

Iwasawa 2007 [35]	%LAA-950, MLD	%LAA-950 & FEV_1: -0.661 %LAA-950 & FEV_1/FVC: -0.745 MLD & FEV_1: 0.636 MLD & FEV_1/FVC: 0.782	NA
Iwasawa 2011 [31]*	%LAA-950	%LAA-950 & FEV_1 %pred: -0.43 %LAA-950 & FEV_1/FVC: -0.49	NA
Jin 2007 [32]	%LAA-960, %LAA-950, %LAA-910, %LAA-900, etc.	%LAA-960 & FEV_1 %pred: -0.501 %LAA-960 & FEV_1/FVC: -0.465 %LAA-950 & FEV_1 %pred: -0.534 %LAA-950 & FEV_1/FVC: -0.513 %LAA-910 & FEV_1 %pred: -0.516 %LAA-910 & FEV_1/FVC: -0.584 %LAA-900 & FEV_1 %pred: -0.470 %LAA-900 & FEV_1/FVC: -0.563	%LAA-910 & FEV_1 %pred: -0.562 %LAA-910 & FEV_1/FVC: -0.506 %LAA-900 & FEV_1 %pred: -0.571 %LAA-900 & FEV_1/FVC: -0.523
Jogi 2011 [59]	Emphysema percentage	No CT quantitative measurements	NA
Kim 2009 [60]	%LAA-950, WT	%LAA-950 & FEV_1 %pred: -0.07 WT & FEV_1 %pred: -0.12	NA
Kosciuch 2009 [11]	WA%, Ai, WA	No reports on the correlation with FEV_1 %pred or FEV_1/FVC in COPD subgroup	NA
Lamers 1994 [8]	Visual score	No CT quantitative measurements	No CT measurements
Leader 2008 [21]*	WA%, Ai, Ao, WA	WA% & FEV_1 %pred: -0.584 Ai & FEV_1 %pred: 0.540 Ao & FEV_1 %pred: 0.410 WA & FEV_1 %pred: 0.172	NA
Leader 2009 [61]	WA%, Ai, Ao, WA, etc.	WA% & FEV_1 %pred: -0.238 WA% & FEV_1/FVC: -0.180 Ai & FEV_1 %pred: 0.286 Ai & FEV_1/FVC: 0.237 Ao & FEV_1 %pred: 0.149 Ao & FEV_1/FVC: 0.128 WA & FEV_1 %pred: -0.007 WA & FEV_1 %pred: -0.001	NA
Lee 2008 [22]*	%LAA-950, MLD, Ai, WA, WA%	%LAA-950 & FEV_1 %pred: -0.547 MLD & FEV_1 %pred: 0.439 WA% & FEV_1 %pred: -0.044	%LAA-950 & FEV_1 %pred: -0.553 MLD & FEV_1 %pred: 0.619
Lee 2011 [62]	%LAA-950	No reports on the correlation with FEV_1 %pred or FEV_1/FVC	No reports on the correlation with FEV_1 %pred or FEV_1/FVC
Lee 2011 [63]	%LAA-950, MLD, LV, etc.	No reports on the correlation with FEV_1 %pred or FEV_1/FVC	No reports on the correlation with FEV_1 %pred or FEV_1/FVC
Li 2009 [64]	LV, etc.	LV & FEV_1 %pred: 0.315 LV & FEV_1/FVC: 0.191	LV & FEV_1 %pred: -0.616 LV & FEV_1/FVC: -0.543
Madani 2010 [65]	%LAA-980 to -900 Perc 1 to 18	No reports on the correlation with FEV_1 %pred or FEV_1/FVC	
Marquez-Martin 2011 [66]	Visual score	No CT quantitative measurements	NA
Matsuoda 2007 [14]	%LAA-950, %LAA-900	%LAA-950 & FEV_1 %pred: -0.471 %LAA-950 & FEV_1/FVC: -0.428 %LAA-900 & FEV_1 %pred: -0.404 %LAA-900 & FEV_1/FVC: -0.320	%LAA-950 & FEV_1 %pred: -0.602 %LAA-950 & FEV_1/FVC: -0.554 %LAA-900 & FEV_1 %pred: -0.618 %LAA-900 & FEV_1/FVC: -0.525
Matsuoda 2008 [67]	Ai, etc.	Ai & FEV_1 %pred: 0.26 Ai & FEV_1/FVC: 0.28	Ai & FEV_1 %pred: 0.63 Ai & FEV_1/FVC: 0.64
Matsuoda 2008 [68]	Percentage between two thresholds	No CT quantitative measurements	No CT quantitative measurements
Mets 2011 [69]	log%LAA-950, Perc15,etc.	log%LAA-950 & logFEV_1: 0.53† log%LAA-950 & FEV_1/FVC: 0.61† Perc15 & logFEV_1: 0.44† Perc15 & FEV_1/FVC: 0.53†	NA
Mishima 1999 [70]	%LAA-960	%LAA-960 & FEV_1 %pred: -0.320 %LAA-960 & FEV_1/FVC: -0.528	NA
Mohamed Hoesein 2011 [45]	%LAA-950, Perc15	%LAA-950 & FEV_1 %pred: -0.16 %LAA-950 & FEV_1/FVC: -0.42 Perc15 & FEV_1 %pred: 0.12 Perc15 & FEV_1/FVC: 0.39	NA
Moron 2004 [71]	Visual score	No CT quantitative measurements	NA
Moroni 2001 [72]	%LAA-910, MLD	No reports on the correlation with FEV_1 %pred or FEV_1/FVC	%LAA-910 & FEV_1/FVC: -0.78 MLD & FEV_1/FVC: 0.85
Nakano 1999 [73]	%LAA-960	%LAA-960 & FEV_1: -0.492 %LAA-960 & FEV_1/FVC: -0.622	NA
Nakano 2000 [3]	%LAA-960 WA%, Ai, Ao, T/D ratio, etc.	%LAA-960 & FEV_1 %pred: -0.529 %LAA-960 & FEV_1/FVC: -0.650 WA% & FEV_1 %pred: -0.338 Ai & FEV_1 %pred: 0.273 Ao & FEV_1 %pred: 0.195 WA% & FEV_1/FVC -0.192	NA
O'Donnel 2004 [40]	%LAA-950, MLD	%LAA-950 & FEV_1 %pred: -0.45 MLD & FEV_1 %pred: 0.38	%LAA-950 & FEV_1 %pred: -0.52 MLD & FEV_1 %pred: 0.63

Ohara 2006 [74]	%LAA-960, WA%, Ai, Ao, T/D ratio, etc.	Separately reported by lung fields %LAA-960 & FEV_1 %pred: -0.331 (upper) %LAA-960 & FEV_1/FVC: -0.222 (upper) %LAA-960 & FEV_1 %pred: -0.487 (lower) %LAA-960 & FEV_1/FVC: -0.491 (lower) WA% & FEV_1 %pred: -0.336 (upper) WA% & FEV_1/FVC: -0.280 (upper) WA% & FEV_1 %pred: -0.339 (lower) WA% & FEV_1/FVC: -0.357 (lower)	NA
Ohno 2011 [30]*	WA%, etc.	WA% & FEV_1 %pred: -0.69 WA% & FEV_1/FVC: -0.59	NA
Orlandi 2004 [15]	%LAA-950, -910, MLD	%LAA-950 & FEV_1 %pred: -0.59 %LAA-950 & FEV_1/FVC: -0.65 Separately reported by radiation dose %LAA-910 & FEV_1 %pred: -0.91 (normal) %LAA-910 & FEV_1/FVC: -0.81 (normal) MLD & FEV_1 %pred: 0.77 (normal) MLD & FEV_1/FVC: 0.78 (normal) %LAA-910 & FEV_1 %pred: -0.87 (low) %LAA-910 & FEV_1/FVC: -0.81 (low) MLD & FEV_1 %pred: 0.57 (low) MLD & FEV_1/FVC: -0.36 (low)	NA
Orlandi 2005 [75]	%LAA-950, MLD, WA%, WA, T/D ratio	%LAA-950 & FEV_1 %pred: -0.42 %LAA-950 & FEV_1/FVC: -0.50 MLD & FEV_1 %pred: 0.40 MLD & FEV_1/FVC: 0.59 WA% & FEV_1 %pred: -0.04 WA% & FEV_1/FVC: -0.009	NA
Park 2008 [23]*	%LAA-950	%LAA-950 & FEV_1 %pred: -0.46 %LAA-950 & FEV_1/FVC: -0.67	NA
Patel 2008 [2]	%LAA-950, %LAA-910, Visual score	%LAA-950 & FEV_1 %pred: -0.31 %LAA-950 & FEV_1/FVC: -0.41 No reports on the correlation between %LAA-910 & FEV_1 %pred or FEV_1/FVC	NA
Pauls 2010 [24]*	%LAA-950, LV	%LAA-950 & FEV_1 %pred: -0.360 LV & FEV_1 %pred: -0.162	NA
Pescarolo 2008 [76]	Visual score	No CT quantitative measurements	NA
Petersen 2010 [77]	WA%, Ai, Ao	No reports on the correlation with FEV_1 %pred or FEV_1/FVC in COPD patients	NA
Sandek 2002 [78]	%LAA-910, MLD	%LAA-910 & FEV_1 %pred: -0.69 %LAA-910 & FEV_1/FVC: -0.76 MLD & FEV_1 %pred: 0.36 MLD & FEV_1/FVC: 0.54	%LAA-910 & FEV_1 %pred: -0.83 %LAA-910 & FEV_1/FVC: -0.87
Scichilone 2008 [79]	MLD, etc	No reports on the correlation between MLD & FEV_1 %pred or FEV_1/FVC	NS
Shaker 2005 [37]	%LAA-910, Perc15, etc.	%LAA-910 & FEV_1 %pred: -0.62 %LAA-910 & FEV_1/FVC: -0.62 Perc15 & FEV_1 %pred: 0.62 Perc15 & FEV_1/FVC: 0.61	NA
Sorensen 2010 [80]	Emphysema pattern	No CT quantitative measurements	NA
Spiropoulos 2003 [81]	%LAA-910	No reports on the correlation with FEV_1 %pred or FEV_1/FVC	No reports on the correlation to FEV_1 %pred or FEV_1/FVC
Torres 2011 [82]	%LAA-960	%LAA-960 & FEV_1 %pred: 0.14 %LAA-960 & FEV_1/FVC: -0.24	NA
Tsushima 2010 [83]	%LAA-960, Visual score	%LAA-960 & FEV_1/FVC: -0.29	NA
Van Der Lee 2006 [84]	%LAA-950, LV	%LAA-950 & FEV_1 %pred: 0.3[†] No reports on the correlation between LV & FEV_1 %pred or FEV_1/FVC	NA
Washko 2008 [12]	%LAA-950, -910, MLD, Perc15, etc	%LAA-950 & FEV_1 %pred: -0.09 %LAA-950 & FEV_1/FVC: -0.09 %LAA-910 & FEV_1 %pred: -0.20 %LAA-910 & FEV_1/FVC: -0.19 MLD & FEV_1 %pred: 0.18 MLD & FEV_1/FVC: 0.21 Perc15 & FEV_1 %pred: 0.09 Perc15 & FEV_1/FVC: 0.12	NA
Washko 2009 [25]*	%LAA-950, WA%, Ai, WT	WA% & FEV_1 %pred: -0.28 WA% & FEV_1/FVC: -0.014 Ai & FEV_1 %pred: 0.14 Ai & FEV_1FVC: 0.07 WT & FEV_1 %pred: -0.13 WT & FEV_1/FVC: -0.05	NA
Watanuki 1994 [85]	MLD	MLD & FEV_1 %pred: 0.72	NA

Yamashiro 2010 [26]*	%LAA-950, MLD, LV, etc.	%LAA-950 & FEV_1 %pred: -0.625 %LAA-950 & FEV_1/FVC: -0.713 MLD & FEV_1.%pred: 0.494 MLD & FEV_1/FVC: 0.562 LV & FEV_1 %pred: -0.010 LV & FEV_1/FVC: -0.198	%LAA-950 & FEV_1 %pred: -0.637 %LAA-950 & FEV_1/FVC: -0.729 MLD & FEV_1.%pred: 0.661 MLD & FEV_1/FVC: 0.743 LV & FEV_1 %pred: -0.406 LV & FEV_1/FVC: -0.588
Yamashiro 2010 [28]*	%LAA-950, WA%, Ai, etc.	%LAA-950 & FEV_1 %pred: -0.460 WA% & FEV_1 %pred: -0.470 Ai & FEV_1 %pred: 0.450	NA
Yamashiro 2011 [86]	Kurtosis and skewness	No CT quantitative measurements	No CT quantitative measurements
Zampatori 1997 [87]	%LAA-900, Visual score	No reports on the correlation with FEV_1 %pred or FEV_1/FVC	%LAA-900 & FEV_1 %pred: -0.65
Zampatori 2001 [88]	Visual score	No CT measurements	NA
Zampatori 2001 [89]	%LAA-900, LV, Visual score	LV & FEV_1 %pred: -0.49 LV & FEV_1/FVC: -0.69	NA
Zampatori 2002 [90]	%LAA-900, MLD, LV, Visual score	%LAA-900 & FEV_1%pred: -0.53(Scan 1) %LAA-900 & FEV_1%pred: -0.56(Scan 2) %LAA-900 & FEV_1/FVC: -0.79 (Scan 1) %LAA-900 & FEV_1/FVC: -0.80 (Scan 2) No reports on the correlation between MLD, LV & FEV_1 %pred or FEV_1/FVC	NA
Zaporozhan 2005 [41]	%-LAA-950, MLD, LV, etc	No reports on the correlation with FEV_1 %pred or FEV_1/FVC	No reports on the correlation to FEV_1 %pred or FEV_1/FVC
Zhang 2008 [27]*	%-LAA-950, -910, MLD	%LAA-950 & FEV_1 %pred: -0.520 %LAA-950 & FEV_1/FVC: -0.626 MLD & FEV_1 %pred: 0.416 MLD & FEV_1/FVC: 0.512 %LAA-910 & FEV_1 %pred: -0.437 %LAA-910 & FEV_1/FVC: -0.548	NA

NA = not available; %LAA = percentage low attenuation area; MLD = mean lung density; LV = lung volume; Perc = percentile point of lung density; WA% = airway wall area percentage; Ai = airway lumen area; Ao = total airway area; WT = wall thickness; T/D radio = ratio of airway wall thickness to total diameter.

Supplemental Table D Assessment of Study Quality, by the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool

Study, Year	Item1: Representative patient sample	Item2: Selection criteria clearly described	Item3: Acceptable reference standard*	Item4: Acceptable delay between tests†	Item5: Partial verification avoided	Item6: Differential verification avoided	Item7: Incorporation avoided	Item8: Adequate index test description*	Item9:Adequate reference standard description*	Item10: Index test blinded to reference standard*	Item11: Reference standard blinded to index test*	Item12: Clinical data available as in practice	Item13: Uninterpretable test results reported	Item14: Withdrawals explained	Score†
Akira 2009 [13]	Y	Y	Y	Y	Y	Y	Y	Y	Y	U	U	Y	Y	Y	13.0
Bon 2009 [18]	Y	Y	Y	U	Y	Y	Y	Y	Y	U	Y	Y	Y	Y	13.0
Dransfield 2007 [19]	Y	Y	Y	U	Y	Y	Y	Y	Y	U	U	Y	Y	Y	12.5
Hasegawa 2006 [20]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	Y	13.5
Hesselbacher 2011[29]	Y	Y	Y	U	Y	Y	Y	Y	Y	U	Y	Y	Y	Y	13.0
Iwasawa 2011 [31]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	Y	13.5
Leader 2008 [21]	Y	Y	Y	U	Y	Y	Y	Y	Y	U	U	Y	Y	Y	12.5
Lee 2008 [22]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	Y	13.5
Ohno 2011 [30]	Y	Y	Y	Y	Y	Y	Y	Y	Y	U	U	Y	Y	Y	13.0
Park 2008 [23]	Y	Y	Y	Y	Y	Y	Y	Y	Y	U	U	Y	Y	Y	13.0
Pauls 2010 [24]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	Y	13.5
Washko 2009 [25]	Y	Y	Y	U	Y	Y	Y	Y	Y	Y	U	Y	Y	Y	13.0
Yamashiro 2010 [26]	Y	Y	Y	U	Y	Y	Y	Y	Y	U	U	Y	Y	Y	12.5
Yamashiro 2010 [28]	Y	Y	Y	U	Y	Y	Y	Y	Y	Y	U	Y	Y	Y	13.0
Zhang 2008 [27]	Y	Y	Y	Y	Y	Y	Y	Y	Y	U	U	Y	Y	Y	13.0
Item number (Yes)	15	15	15	8	15	15	15	15	15	6	2	15	15	15	
Item number (No)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Item number (Un-known)	0	0	0	7	0	0	0	0	0	9	13	0	0	0	

Y = yes; U = unclear; N = no.

Supplemental Table E Bias analysis in the meta-analysis

	Heterogeneity			Publication bias			Meta-regression for gender*
	Q value	p(Q)	I^2,%	Fail-safe N	Begg and Mazumdar rank correlation, p	Egger's regression, p	p(Slope)
Inspiration							
%LAA-950 and FEV_1%pred	11.65	0.11	39.9	451	0.62	0.22	0.16
%LAA-950 and FEV_1/FVC	10.22	0.12	41.3	614	0.45	0.31	0.71
MLD and FEV_1%pred	6.08	0.11	50.6	97	0.50	0.14	0.22
MLD and FEV_1/FVC	8.29	0.02	75.9	172	0.60	0.77	0.05
WA% and FEV_1%pred	8.50	0.04	64.7	255	0.73	0.60	IS
Expiration							
%LAA-950 and FEV_1%pred	0.81	0.67	0	84	0.12	0.21	0.85
%LAA-950 and FEV_1/FVC	0.38	0.54	0	IS	IS	IS	IS
MLD and FEV_1%pred	3.55	0.17	43.6	122	0.12	0.15	0.56
MLD and FEV_1/FVC	0.05	0.82	0	IS	IS	IS	IS

FEV_1 %pred = predicted forced expiratory volume in the first second; FEV_1/FVC = FEV_1 divided by forced vital capacity; %LAA-950 = percentage lower attenuation area than -950 HU; MLD = mean lung density; Perc 15 = 15 percentile point of lung density; WA% = wall area percentage; IS = insufficient study numbers to perform analysis.

Chapter 9

Chronic Respiratory Symptoms Associate with Airway Wall Thickening Measured by Thin-Slice Low-Dose CT

Submitted

Xueqian Xie[12] / Akkelies E. Dijkstra[3] / Judith M. Vonk[4]
Matthijs Oudkerk[2] / Rozemarijn Vliegenthart[12] / Harry J.M. Groen[3]

University of Groningen, University Medical Center Groningen,
[1]Department of Radiology, [2]Center for Medical Imaging – North East Netherlands,
[3]Department of Pulmonary Diseases, [4]Department of Epidemiology

Abstract

Purpose

The prevalence of chronic respiratory symptoms is high in smokers in lung cancer screening. The study-purpose is to compare CT-derived airway wall measurements between smokers with and without chronic respiratory symptoms.

Materials and Methods

50 heavy smokers with chronic respiratory symptoms (cough, mucus, dyspnea and wheezing) and 50 without any respiratory symptom were randomly selected from the NELSON trial. This trial was ethically approved. All participants gave written informed consent. Images on thin-slice low-dose CT were evaluated using dedicated software for airway measurements. Wall area percentage (WA%) and airway wall thickness (AWT) and were measured for airways with a luminal diameter ≥ 5 mm in five different bronchi in different pulmonary lobes. Association between airway wall measurements and respiratory symptoms was analyzed using multiple linear regression, adjusted for age, body mass index, smoking status, emphysema and pulmonary function.

Results

After adjusting for relevant factors, a significant positive association between airway wall measurements and respiratory symptoms was found in airways with a 5 - 10 mm luminal diameter ($p < 0.01$), but not in ≥ 10 mm airways ($p > 0.05$). In the 5 - 10 mm airway level, mean WA% was 51.5 ± 7.9% and 48.1 ± 7.7% in symptomatic and asymptomatic group, respectively. AWT was 1.54 ± 0.39 mm and 1.37 ± 0.35 mm, respectively.

Conclusion

Heavy smokers with chronic respiratory symptoms in lung cancer screening, who are of high-risk for chronic bronchitis, have bronchial wall thickening in airways with luminal diameter from 5 to 10 mm, but not in larger airways.

Introduction

Nearly half of smokers in lung cancer screening have chronic respiratory symptoms, i.e., chronic hyper secretion of mucus, combined with chronic cough, often accompanied by dyspnea and wheezing [1]. Smokers with these symptoms have a great chance to develop chronic bronchitis, a disease associated with an accelerated decline in pulmonary function – a major risk factor for chronic obstructive pulmonary disease (COPD) and all-cause mortality [2,3]. During their lifetime, over 40% of smokers develop chronic bronchitis [2]. The clinical diagnosis of chronic bronchitis is commonly based on a combination of medical history, physical examination, spirometry and laboratory test [4]. The confirmative diagnosis relies on histopathology [5]. Despite the high prevalence, chronic bronchitis was often under-diagnosed or late-diagnosed [6].

Chronic bronchitis is histopathologically found in a range of airways, commonly in large airways [7]. The morphological basis of chronic bronchitis is bronchial wall thickening and airway luminal narrowing, which subsequently result in airflow limitation [8]. Morphological changes are important to understand pathogenesis and effect of therapeutic interventions for chronic bronchitis [5,9]. Recent development of thin-slice multi-detector CT and dedicated software, allows accurate noninvasive quantification of airway dimensions in large bronchi [10-12].

In participants of lung cancer screening, it is important to know whether there are morphological changes of airway walls in subjects with chronic respiratory symptoms, since there would be substantial benefit to early detect airway remodelling of chronic bronchitis in this high-risk population. Airway wall thickening associated with some relevant factors, such as age, body mass index, smoking status, emphysema and pulmonary function [13,14]. In the population of lung cancer screening, the adjusted association between CT-derived airway wall quantification and respiratory symptoms is still unclear. Therefore, the purpose of our study was to retrospectively compare the airway wall thickness along the respiratory pathway between subjects with and without chronic respiratory symptoms, adjusted for relevant factors.

Materials and methods

Sample

The study sample was randomly selected from the baseline round in one center of a population-based Dutch-Belgian Randomized Lung Cancer Screening Trial (NELSON) (Figure 1). The symptomatic group contained 50 heavy smokers with four chronic respiratory symptoms, including chronic cough, chronic mucus hyper-secretion, dyspnea and wheezing lasting for at least three months during the last year before inclusion. The asymptomatic

group contained 50 smokers without any respiratory symptoms. Given the confidence level of 95%, a 50/50 sample size would provide a statistical power of > 0.90 in t-test when assessing wall area percentage (WA%) in large airways between symptomatic and asymptomatic groups [15].

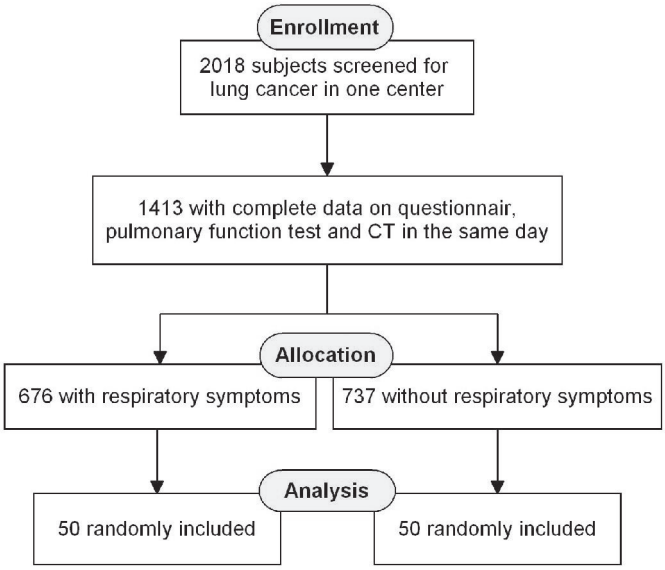


Figure 1 Flow diagram of subject selection

The NESON trial was approved by the Dutch Minister of Health and the ethics board in the participating center. All participants gave written informed consent. Detailed inclusion and exclusion criteria and characteristics in this male population-based trial have been described elsewhere [16]. In short, current or former elderly smokers with a smoking history of at least 20 pack-years were included. Individuals in moderate or poor health with inability to climb two flights of stairs were excluded. Information about the presence of respiratory symptoms and smoking behavior (current or former smoker, pack-years, etc.) was obtained by questionnaires.

CT imaging

A 16-row multi-detector CT (Sensation 16, Siemens, Forchheim, Germany) was utilized with a low-dose acquisition protocol [17]. The protocol was: spiral acquisition at 120 kV, 20 mAs, rotation time 0.5 s, pitch 1.5 and collimation 16 × 0.6 mm, field of view 350 mm, slice thickness 1 mm and slice increment 0.7 mm. The effective radiation dose is less than 0.8 mSv. Contrast media was not used. The images were reconstructed to a pixel matrix of 512 × 512 using a medium-smooth B30f kernel. The CT system was calibrated routinely. The image acquisition was performed during one breath-holding at full inspiration after appropriate instruction.

Pulmonary function testing

Standard pulmonary function testing was performed according to the European Respiratory Society guidelines [18] on the same day as the CT acquisition. In this population based trial, a bronchodilator was not administered. Forced expiratory volume in one second (FEV_1) and forced vital capacity (FVC) were assessed. FEV_1 was presented as a percentage of predicted (FEV_1 %pred).

Airway selection

We investigated large airways with an internal luminal diameter ≥ 5 mm [19, 20], which were categorized into three categories (luminal diameter $5 \leq Ø < 10$ mm; $10 \leq Ø < 15$ mm; and $Ø \geq 15$ mm). To measure the airway dimensions, we selected five bronchi of segmental level, each representing one pulmonary lobe: the apical bronchus (RB1) of the right upper lobe, the lateral bronchus (RB4) of the right middle lobe, the posterior basal bronchus (RB10) of the right lower lobe, the apicoposterior bronchus (LB1+2) of the left upper lobe, and posterior basal bronchus (LB10) of the left lower lobe. Those bronchi were selected because they are relatively free from cardiac motion artifacts. In the bronchial tree below segmental level, when there was more than one bronchus in each airway generation, we evaluated only one bronchus in each generation. Thereafter, the bronchial pathway was evaluated from trachea down to airway level of 5 mm luminal diameter.

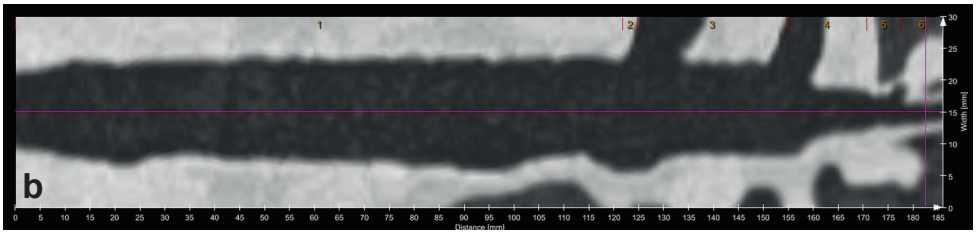


Figure 2 Reformatted images generated by the dedicated software. The bronchial tree was automatically segmented (**a**). A bronchial pathway from trachea to a selected bronchus was converted to a stretched multiple planar reconstruction (MPR) image (**b**).

Quantitative image analysis

The images were evaluated using dedicated software for airway measurement (Airway Examiner 1.0, Fraunhofer MEVIS, Bremen, Germany), which based on a three-dimensional (3D) algorithm of airway geometry [12], instead of the traditional full-width-at-half-maximum (FWHM) algorithm. Briefly, wall thickness is estimated in 3D space using this 3D algorithm, rather than 2D plane in FWHM method [12]. For airway wall thickness as small as 1 mm, the 3D algorithm showed much better accuracy and reproducibility than the FWHW method in phantom studies [12,21]. This software tool follows two principal steps. Firstly, the software automatically segments bronchial tree. Secondly, after clicking a bronchus in bronchial tree, the software automatically quantifies airway dimensions along trachea to the chosen bronchus, presenting data per 1 mm along this respiratory pathway. A representative figure generated by the software package was shown in Figure 2. We collected three airway quantitative measurements, including WA%, airway wall thickness (AWT) and airway luminal diameter. WA% was defined as (wall area) / (wall area + lumen area) × 100%.

After randomization of all included subjects, one radiologist with nine years' experiences in thoracic diagnostic radiology evaluated the images, blinded to subject information on basic characteristics and respiratory symptoms during evaluation. Time duration of whole assessment processes for each subject was approximately three minutes.

Emphysema quantification was performed using dedicated software (ImageXplorer, Image Sciences Institute, Utrecht, The Netherlands). This software automatically implies lung segmentation, image noise reduction, and CT density calibration to improve accuracy and reproducibility of evaluation [22,23]. The images were recalibrated by shifting CT density so that the density inside the trachea became -1000 Hounsfield unit (HU) [19]. We collected three CT quantification parameters, including 15 percentile point of lung density (Perc 15), percentage of lung attenuation area under -950 HU (%LAA-950) and lung volume. Larger emphysematous tissues are indicated by lower Perc 15 or higher %LAA-950.

Statistics

Data are reported as mean ± standard deviation (SD) for normally distributed data or median (25^{th}, 75^{th} percentile) for nonnormally distributed data. Differences in characteristics between the symptomatic and asymptomatic group were assessed by independent-samples t-test for normally distributed continuous data, by Mann-Whitney U-tests for nonnormally distributed continuous data, and by Chi-square test for nominal data. The association between airway wall measurements (WA% and AWT) and potentially associated factors was evaluated by univariate linear regression. The association between airway wall measurements and respiratory symptoms was analyzed using multiple linear regression, adjusted for age, BMI, smoking status, Perc 15, and FEV_1 %pred.

A $p < 0.05$ was considered as statistically significant. Statistical analyses were performed using SPSS version 20 (IBM, New York, US).

Results

Sample characteristics

Characteristics of the symptomatic and asymptomatic group are presented in Table 1. All subjects were male, with a mean age of 56.5 ± 5.4 years (range from 50 to 69 years). No diseases (pneumonia, atelectasis and heavy pulmonary fibrosis, etc.) affecting airway wall measurements were observed through reviewing CT images. No obstructive diseases (airway mass/tumor, external compression and bronchial stricture, etc.) causing airway wall thickening were found either. Airway walls were successfully evaluated in all the 100 subjects, in which, 65,070 cross sections of airways were measured. WA% increased from proximal to distal airway, while AWT and luminal diameter decreased. WA% ranged from

14.6 to 75.5%. AWT ranged from 0.7 to 3.2 mm. Airway luminal diameter ranged from 5.0 to 22.9 mm.

Table 1 Characteristics of included subjects with and without chronic respiratory symptoms

	Symptomatic	Asymptomatic	p value
Sample size, *n*	50	50	
Basic characteristics			
Male, *n*	50	50	
Age, years	56.0 ± 5.1	57.3 ± 5.7	0.190
Weight, kg	86.9 ± 13.6	83.9 ± 12.5	0.253
Height, m	1.79 ± 0.05	1.79 ± 0.06	0.628
Body mass index, kg/m^2	27.0 ± 3.9	26.2 ± 3.3	0.282
Smoking behavior			
Current/former smoker, *n*	37 / 13	23 / 27	< 0.05
Pack-years	42.1 ± 15.5	39.9 ± 14.6	0.411
Smoking duration, years	8.3 ± 1.0	8.0 ± 1.1	0.130
CT quantification			
Wall thickness, mm	1.55 ± 0.44	1.42 ± 0.40	< 0.001
Wall area percentage, %	47.0 ± 12.1	43.3 ± 11.1	< 0.001
Lung volume, l	7.0 (6.3, 8.0)	6.6 (5.7, 7.5)	< 0.001
Perc15, HU	-922 (-933, -912)	-915 (-928, -898)	< 0.001
%LAA-950, %	3.2 (2.1, 5.5)	2.3 (1.0, 4.0)	< 0.001
Pulmonary function test			
FEV_1 %pred	80.3 (65.4, 110.4)	102.9 (95.6, 108.2)	< 0.001
FEV_1/FVC, %	65.5 (51.9, 72.4)	75.3 (69.6, 80.8)	< 0.001
COPD status			
COPD, *n*	34	13	< 0.001
GOLD stage I, *n*	12	11	
GOLD stage II, *n*	14	2	
GOLD stage III, *n*	6	0	
GOLD stage IV, *n*	0	0	

Data are reported as mean ± standard deviation for normally distributed data or median (25^{th}, 75^{th} percentile) for nonnormally distributed data. HU = Hounsfield unit; Perc15 = 15 percentile point of lung density; %LAA-950 = percentage of lung attenuation area under -950HU; FEV_1 %pred = forced expiratory volume in the first second as percentage from predicted; FEV_1/FVC = FEV_1 divided by forced vital capacity; COPD = chronic obstructive pulmonary disease; GOLD = the Global Initiative for Chronic Obstructive Lung Disease.

The symptomatic group had more current smokers than asymptomatic group (74% vs. 46%, $p < 0.05$). No significant differences were found for age, BMI, pack-years and smoking duration between those two groups ($p > 0.05$).

In CT emphysema quantification, the symptomatic group had significantly lower Perc 15 and higher %LAA-950 than the asymptomatic group ($p < 0.001$), indicating more emphysematous tissues in the symptomatic group. Median %LAA-950 was 3.2 (25^{th}, 75^{th} percentile: 2.1, 5.5) and 2.3 (1.0, 4.0) in the symptomatic and asymptomatic group, respectively, indicating mild emphysema in those two groups [10]. The symptomatic group had significantly worse pulmonary function (FEV_1 %pred and FEV_1/FVC) than asymptomatic group ($p < 0.001$).

Table 2 Univariate regression analysis for factors associated with airway wall measurements

	Airway wall thickness		Wall area percentage	
	B	p value	B	p value
Basic characteristics				
Age, years	-0.003	< 0.05	-0.132	< 0.01
Weight, kg	0	0.508	0.041	0.281
Height, m	0.151	0.373	1.940	0.667
Body mass index, kg/m^2	0.012	< 0.05	0.160	< 0.05
Smoking behavior				
Current smoking	0.102	< 0.001	2.418	< 0.001
Pack-years	0.004	< 0.001	0.084	< 0.001
Smoking duration, years	0.021	< 0.001	0.233	< 0.01
Respiratory symptoms				
Presence of respiratory symptoms	0.124	< 0.001	3.742	< 0.001
CT emphysema quantification				
Lung volume, l	0.021	0.244	-0.163	0.361
Perc15, HU	0	0.259	0.021	0.132
%LAA-950, %	0.002	0.242	-0.041	0.430
Pulmonary function test				
FEV_1 %pred	-0.005	< 0.001	-0.120	< 0.001
FEV_1/FVC, %	-0.007	< 0.001	-0.164	< 0.001

HU = Hounsfield unit; Perc 15 = 15 percentile point of lung density; %LAA-950 = percentage of lung attenuation area under -950 HU; FEV_1 %pred = forced expiratory volume in the first second as percentage from predicted; FEV_1/FVC = FEV_1 divided by forced vital capacity.

Factors associated with airway wall measurements

Univariate linear regression analysis showed that airway wall measurements positively associated with the presence of respiratory symptoms, and with BMI, current smoker, pack-years and smoking duration ($p < 0.05$) (Table 2). Conversely, airway wall measurements negatively associated with age and pulmonary function parameters (FEV_1 %pred and FEV_1/FVC) ($p < 0.001$).

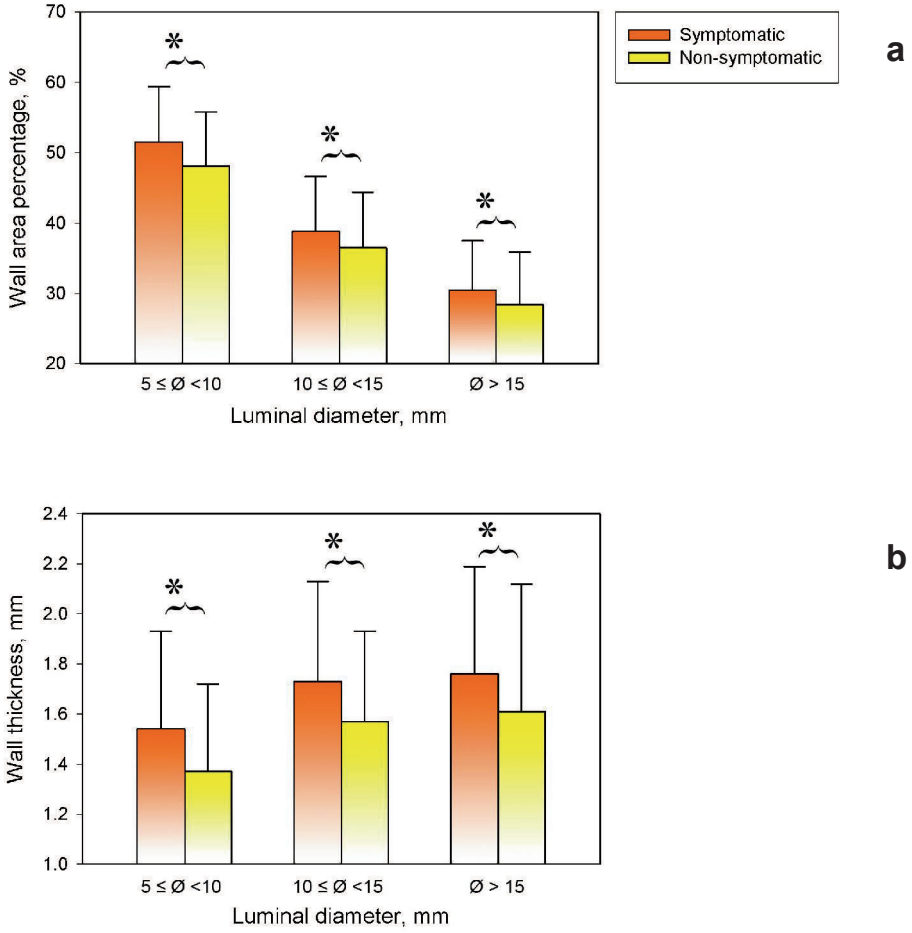


Figure 3 Airway wall area percentage (**a**) and thickness (**b**) between symptomatic and asymptomatic subjects of chronic respiratory symptoms, without adjusting for relevant factors. Significant difference between symptomatic and asymptomatic group was indicated as "*".

Airway wall measurements between symptomatic and asymptomatic group

Without adjustment for relevant factors, the symptomatic group showed overall higher WA% (47.0 ± 12.1% vs. 43.3 ± 11.1%, $p < 0.001$) and higher AWT (1.55 ± 0.44 mm vs. 1.42 ± 0.40 mm, $p < 0.001$) than the asymptomatic group. In detail, in all the three categories of large airways (luminal diameter: $5 \leq Ø < 10$, $10 \leq Ø < 15$ and $Ø \geq 15$), the symptomatic group showed significant higher WA% and AWT ($p < 0.01$) (Figure 3). The representative images of those two groups are showed in Figure 4.

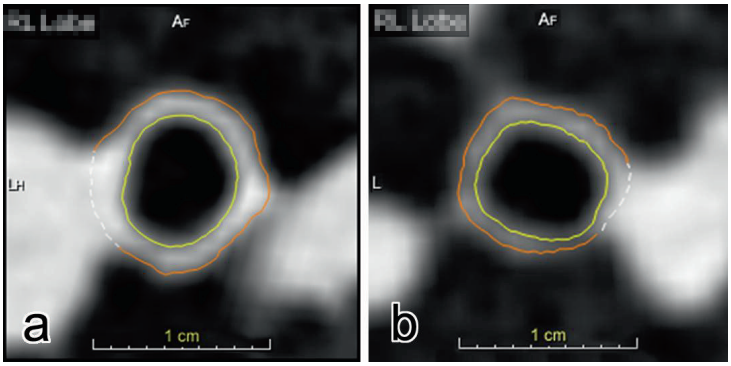


Figure 4 Cross-section images perpendicular to long axis of the bronchi. An image of a symptomatic subject (**a**). Wall area percentage (WA%) = 51%, airway wall thickness (AWT) = 1.4 mm and luminal diameter = 7 mm. An image of an asymptomatic subject (**b**). WA% = 43%, AWT = 1.1 mm and luminal diameter = 7 mm.

However, using multiple linear regression, after adjusted for age, BMI, smoking status, Perc 15, and FEV_1 %pred, a significant positive association between airway wall measurements (WA% and AWT) and respiratory symptoms was found only in airways with a luminal diameter from 5 to 10 mm ($p < 0.01$). In the airway level from 5 to 10 mm, mean WA% was 51.5 ± 7.9% and 48.1 ± 7.7% in symptomatic and asymptomatic group, respectively. AWT was 1.54 ± 0.39 mm and 1.37 ± 0.35 mm, respectively. No significant associations were found in airways with a luminal diameter ≥ 10 mm ($p > 0.05$) (Table 3).

Table 3 Multiple linear regression for the association between airway wall measurements and Chronic respiratory symptoms, adjusted for age, BMI, current smoking status, Perc 15, and FEV_1 %pred

		Airway luminal diameter, mm					
		5 ≤ Ø < 10		10 ≤ Ø < 15		Ø ≥ 15	
		B	p value	B	p value	B	p value
Wall thickness	Respiratory symptoms	0.091	<0.001	0.051	0.170	0.052	0.339
	Age	0.004	0.075	0.003	0.409	-0.013	<0.01
	BMI	0.014	<0.001	0.002	0.646	-0.009	0.186
	Current smoking	0.019	0.461	0.159	<0.001	-0.014	0.793
	Perc 15	0.003	<0.001	0	0.618	-0.001	0.228
	FEV_1 %pred	-0.006	<0.001	-0.004	<0.001	-0.003	<0.01
Wall area percentage	Respiratory symptoms	1.647	<0.01	0.054	0.944	0.551	0.500
	Age	0.042	0.348	-0.065	0.347	-0.164	<0.05
	BMI	0.333	<0.001	0.015	0.881	-0.131	0.219
	Current smoking	0.509	0.330	3.084	<0.001	-0.202	0.801
	Perc 15	0.077	<0.001	-0.017	0.332	-0.019	0.305
	FEV_1 %pred	-0.130	<0.001	-0.076	<0.001	-0.059	<0.001

BMI = body mass index; Perc 15 = 15 percentile point of lung density; FEV_1 %pred = forced expiratory volume in the first second as percentage from predicted.

Discussion

Thin-slice CT and automated software are promising tools to quantify airway walls. Using these techniques, after adjustment for relevant factors, heavy smokers with chronic respiratory symptoms have significant thicker airway walls in airways with luminal diameter from 5 to 10 mm, but not the larger airways. If not adjusted for, the thicker airway walls were in all ≥ 5 mm airways.

The symptomatic group showed general thicker airway walls, up to the trachea. The common causes of bronchial wall thickening are inflammatory, congenital (e.g., cystic fibrosis, alpha-1 antitrypsin deficiency, etc.) and obstructive [24, 25]. Inflammation of the mucous membrane directly results in hyper-secretion of mucus, leading to respiratory symptoms, such as cough, dyspnea and wheezing [26]. Our sample was from a population-based trial, and the prevalence of these congenital bronchial diseases is rare [27, 28]. An experienced radiologist reviewed the CT images, and did not observe obstructive bronchial diseases. Thus, the primary cause of bronchial wall thickening in our study is likely inflammatory. Chronic bronchitis is a disease associated with long-term inflammatory stimulation [7].

Using CT quantification in COPD population instead of heavy smokers, Patel, et al. observed a significant positive association between wall thickness and respiratory symptoms in approximately 6 mm airways in luminal diameter [29]. Mair, et al. found a significant positive association in proximal airways (> 11 mm in luminal diameter), but not for distal airways (approximately 2 to 4 mm) [20]. In consistent with the histological evidence that inflammation and airway remodeling, associating with chronic bronchitis, is located in the more central airways [30], our results strengthened those previous noninvasive findings, in case of without adjustment for relevant factors.

Importantly, we adjusted for five relevant factors to determine the adjusted association between thicker airway walls and chronic respiratory symptoms. In accordance with the previous studies, where age, BMI, smoking status and pulmonary function associated with airway wall thickening [14, 31, 32], we found that thicker airway walls were significantly associated with younger age, higher BMI, current smoking and worse pulmonary function by univariate linear regression. Inconsistent with the prior studies, where emphysema associated with airway wall thickening [14, 33], we found a nonsignificant association between CT emphysema quantification and airway wall thickening. That inconsistent finding might be explained by only mild emphysema in our sample, selected from a population-based screening trial. Our symptomatic group has more current smokers. Smoking often causes more active airway inflammation [34], which is an important potential confounder in this study. After adjustment for these potential confounders, we found signifi-

cant thicker airway walls in airways with luminal diameter from 5 to 10 mm, but not the larger airways.

A recently introduced three-dimensional algorithm was used to assess airway walls in this study. With this algorithm, wall thickness is approximated by an integral based closed-form solution, based on the volume conservation property of convolution [12]. In contrast, the traditionally utilized "full-width-at-half-maximum (FWHM)" algorithm calculates the X-ray attenuation values along rays placed from the lumen center to outward directions in cross section [11]. The former algorithm has shown much better accuracy and reproducibility than the latter, for wall thickness as small as 1 mm in phantom studies [12, 21]. Thus, we expected that our measurements were accurate, in case of the sub-millimeter thickness difference between symptomatic and asymptomatic group.

CT quantification is associated with pathophysiological changes of airway remodeling. A number of CT-derived measurements have been used for airways quantification, such as thickness, area, perimeter, CT density and visual score [13, 29, 35]. We measured WA% and AWT, because those two measurements directly indicate, and pathologically reflect airway wall thickening [36]. Moreover, we utilized automated dedicated software to evaluate airway dimensions per 1 mm. Some prior studies used manual methods to quantify airway walls, where airway walls were noncontinuously measured on cross-sections in-between a large gap of up to 20 mm [15, 29]. Mair, et al. has quantified airway walls, as a function to airway generation [20]. In addition to that study, we evaluated airway walls as a function to airway luminal diameter, since the same airway generation in two bronchi might be of different size.

We evaluated large airways of ≥ 5 mm in luminal diameter, because chronic bronchitis commonly involves large airways. On the other hand, thickening of small airway walls is important for the pathogenesis of COPD and asthma [36, 37]. Increased CT-derived airway thickness of more peripheral airways, as small as 2 to 4 mm in luminal diameter, strongly correlated with airflow limitations in those diseases [19, 38, 39].

Clinical implications

In lung cancer screening trials, the common participants are heavy smokers, with a high prevalence of chronic bronchitis [2, 16]. Despite the high prevalence, chronic bronchitis was often under-diagnosed or late-diagnosed [6]. One morphological manifestation of chronic bronchitis is bronchial wall thickening caused by chronic inflammatory stimulation [7]. We found that heavy smokers with chronic respiratory symptoms had significant thicker airway walls, which represents airway remodeling in an inflammatory process. Our study shows that this airway remodeling can be detected using thin-slice CT and dedicated software, thus has potential benefit for early diagnosis of chronic bronchitis.

Currently, clinical symptoms and spirometry are commonly used for diagnosis and surveillance of chronic bronchitis [40]. Noninvasive CT quantification of airway walls has shown the potential for regional and morphological evaluation of the therapeutic response in treatment of chronic bronchitis [9]. In our study, significantly thicker airway walls are especially identified in large airways of 5 to 10 mm in luminal diameter in respiratory symptomatic group. Thus, CT quantification of airway walls may provide additional morphological information beyond clinical symptoms and spirometry. When assessing airway wall thickening in symptomatic individuals, the airways of 5 to 10 mm in luminal diameter optimally reflect the presence of respiratory symptoms, but not the larger airways.

The absolute increase of airway wall thickening in symptomatic subjects is commonly within a millimeter, thus is difficult to be perceptible to human eyes on CT images. Hence, dedicated software is essential to assess airway walls. Our results were based on 16-row multi-detector CT utilized in a long-standing screening trial, we expect the latest CT technique could provide higher spatial resolution thus improve airway wall assessment. For example, we used image matrix of 512 × 512 voxels, while 1024 × 1024 matrix has been available [41]. Thin-slice CT and automated software showed promising results in this study, it is interesting to know the strength of these techniques in diagnosis of chronic bronchitis.

Limitations

Firstly, inherent to a population-based lung cancer screening trial, histopathological results of bronchial wall are very hard to be available. At least, our results suggested that the chronic respiratory symptoms associated with airway remodeling in an inflammatory process, which is a pathological basis for chronic bronchitis. Also, participants are only heavy male smokers with mild emphysema. Gender was not adjusted for in this study. We expected our included sample at least represent the common population in lung cancer screening. Secondly, a post-bronchodilator pulmonary function test was not performed to assess reversibility in airflow limitation, which is a criterion to exclude bronchial asthma [37]. However, our results might not be substantially influenced by that limitation. Asthma, whether or not coexisting with chronic bronchitis, generally involves only a small number of elderly heavy smokers [42]. Our sample was from a population-based trial, and the prevalence of asthma in elderly men is approximately 2 % in the Netherlands [43]. Thirdly, five bronchi from different pulmonary lobes were evaluated because they are relatively free from cardiac motion artifacts, instead of bronchi from all the pulmonary lobes. A large number of airway cross sections (650 per subject) were measured. We expected that five bronchi with a large number of measurements could represent the quantification of airway dimensions.

Conclusions

After adjustment for relevant factors, heavy smokers with chronic respiratory symptoms out of a population-based lung cancer screening trial have significantly thicker airway walls than asymptomatic smokers. Thus, heavy smokers with chronic respiratory symptoms do have airway remodeling. Thin-slice CT and dedicated software showed the potential to evaluate airway remodeling in smokers with chronic respiratory symptoms.

References

1. Mets OM, Buckens CF, Zanen P, et al. (2011) Identification of chronic obstructive pulmonary disease in lung cancer screening computed tomographic scans. JAMA: the journal of the American Medical Association 306:1775-1781
2. Pelkonen M (2008) Smoking: relationship to chronic bronchitis, chronic obstructive pulmonary disease and mortality. Curr Opin Pulm Med 14:105-109
3. Guerra S, Sherrill DL, Venker C, Ceccato CM, Halonen M, Martinez FD (2009) Chronic bronchitis before age 50 years predicts incident airflow limitation and mortality risk. Thorax 64:894-900
4. Laitinen LA, Koskela K, Expert Advisory G (1999) Chronic bronchitis and chronic obstructive pulmonary disease: Finnish National Guidelines for Prevention and Treatment 1998-2007. Respiratory Medicine 93:297-332
5. Mullen JB, Wright JL, Wiggs BR, Pare PD, Hogg JC (1985) Reassessment of inflammation of airways in chronic bronchitis. Br Med J (Clin Res Ed) 291:1235-1239
6. Huchon GJ, Vergnenegre A, Neukirch F, Brami G, Roche N, Preux PM (2002) Chronic bronchitis among French adults: high prevalence and underdiagnosis. Eur Resp J 20:806-812
7. Hogg JC (2004) Pathophysiology of airflow limitation in chronic obstructive pulmonary disease. Lancet 364:709-721
8. Hogg JC, Timens W (2009) The pathology of chronic obstructive pulmonary disease. Annual review of pathology 4:435-459
9. Hamada S, Ito Y, Imai S, Oguma T, Niimi A, Mishima M (2011) Effect of inhaled corticosteroid therapy on CT scan-estimated airway dimensions in a patient with chronic bronchitis related to ulcerative colitis. Chest 139:930-932
10. Shaker SB, Dirksen A, Bach KS, Mortensen J (2007) Imaging in chronic obstructive pulmonary disease. COPD-Journal of Chronic Obstructive Pulmonary Disease 4:143-161
11. Mets OM, de Jong PA, van Ginneken B, Gietema HA, Lammers JWJ (2012) Quantitative computed tomography in COPD: possibilities and limitations. Lung 190:133-145
12. Weinheimer O, Achenbach T, Bletz C, Dueber C, Kauczor HU, Heussel CP (2008) About objective 3-D analysis of airway geometry in computerized tomography. IEEE Transactions on Medical Imaging 27:64-74
13. Xie X, de Jong PA, Oudkerk M, et al. (2012) Morphological measurements in computed tomography correlate with airflow obstruction in chronic obstructive pulmonary disease: systematic review and meta-analysis. European radiology 22:2085-2093
14. Kim WJ, Silverman EK, Hoffman E, et al. (2009) CT metrics of airway disease and emphysema in severe COPD. Chest 136:396-404

15. Orlandi I, Moroni C, Camiciottoli G, et al. (2005) Chronic obstructive pulmonary disease: Thin-section CT measurement of airway wall thickness and lung attenuation. Radiology 234:604-610
16. Ru Zhao Y, Xie X, de Koning HJ, Mali WP, Vliegenthart R, Oudkerk M (2011) NELSON lung cancer screening study. Cancer imaging 11 S79-84
17. van Klaveren RJ, Oudkerk M, Prokop M, et al. (2009) Management of lung nodules detected by volume CT scanning. N Engl J Med 361:2221-2229
18. Quanjer PH, Tammeling GJ, Cotes JE, Pedersen OF, Peslin R, Yernault JC (1993) Lung volumes and forced ventilatory flows. Report Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. Official Statement of the European Respiratory Society. The European respiratory journal Supplement 16:5-40
19. Dijkstra AE, Postma DS, Ten Hacken N, et al. (2013) Low-dose CT measurements of airway dimensions and emphysema associated with airflow limitation in heavy smokers: a cross sectional study. Respiratory research 14:11
20. Mair G, Maclay J, Miller JJ, et al. (2010) Airway dimensions in COPD: relationships with clinical variables. Respir Med 104:1683-1690
21. Schmidt M, Kuhnigk JM, Krass S, Owsijewitsch M, de Hoop B, Peitgen HO (2010) Reproducibility of airway wall thickness measurements. Proceedings of the SPIE - The International Society for Optical Engineering 7624:762410-762419
22. Schilham AM, van Ginneken B, Gietema H, Prokop M (2006) Local noise weighted filtering for emphysema scoring of low-dose CT images. IEEE Trans Med Imaging 25:451-463
23. van Rikxoort EM, de Hoop B, Viergever MA, Prokop M, van Ginneken B (2009) Automatic lung segmentation from thoracic computed tomography scans using a hybrid approach with error detection. Med Phys 36:2934-2947
24. Hyde DM, King TE, McDermott T, et al. (1992) Idiopathic pulmonary fibrosis - quantitative assessment of lung pathology - comparison of a semiquantitative and a morphometric histopathologic scoring system. Am Rev Respir Dis 146:1042-1047
25. Thurlbeck WM (1990) Pathology of chronic air-flow obstruction. Chest 97:S6-S10
26. Society AT (1995) Standards for the diagnosis and care of patients with chronic obstructive pulmonary disease. . Am J Respir Crit Care Med 152:S77-121
27. Farrell PM (2008) The prevalence of cystic fibrosis in the European Union. Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society 7:450-453
28. Rahaghi FF, Sandhaus RA, Strange C, et al. (2012) The prevalence of alpha-1 antitrypsin deficiency among patients found to have airflow obstruction. Copd 9:352-358
29. Patel BD, Coxson HO, Pillai SG, et al. (2008) Airway wall thickening and emphysema show independent familial aggregation in chronic obstructive pulmonary disease. Am J Respir Crit Care Med 178:500-505

30. Hogg JC, Chu F, Utokaparch S, et al. (2004) The nature of small-airway obstruction in chronic obstructive pulmonary disease. The New England journal of medicine 350:2645-2653
31. Donohue KM, Hoffman EA, Baumhauer H, et al. (2012) Cigarette smoking and airway wall thickness on CT scan in a multi-ethnic cohort: The MESA Lung Study. Respiratory Medicine 106:1655-1664
32. Zach JA, Newell JD, Jr., Schroeder J, et al. (2012) Quantitative computed tomography of the lungs and airways in healthy nonsmoking adults. Investigative radiology 47:596-602
33. Kim Y-I, Schroeder J, Lynch D, et al. (2011) Gender differences of airway dimensions in anatomically matched sites on CT in smokers. COPD-Journal of Chronic Obstructive Pulmonary Disease 8:285-292
34. Battaglia S, Mauad T, van Schadewijk AM, et al. (2007) Differential distribution of inflammatory cells in large and small airways in smokers. J Clin Pathol 60:907-911
35. Yamashiro T, Matsuoka S, Estepar RS, et al. (2010) Quantitative assessment of bronchial wall attenuation with thin-section CT: An indicator of airflow limitation in chronic obstructive pulmonary disease. AJR Am J Roentgenol 195:363-369
36. Aysola RS, Hoffman EA, Gierada D, et al. (2008) Airway remodeling measured by multidetector CT is increased in severe asthma and correlates with pathology. Chest 134:1183-1191
37. Rabe KF, Hurd S, Anzueto A, et al. (2007) Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease - GOLD executive summary. Am J Respir Crit Care Med 176:532-555
38. Matsuoka S, Kurihara Y, Yagihashi K, Hoshino M, Nakajima Y (2008) Airway dimensions at inspiratory and expiratory multisection CT in chronic obstructive pulmonary disease: Correlation with airflow limitation. Radiology 248:1042-1049
39. Hasegawa M, Nasuhara Y, Onodera Y, et al. (2006) Airflow limitation and airway dimensions in chronic obstructive pulmonary disease. Am J Respir Crit Care Med 173:1309-1315
40. Blasi F, Ewig S, Torres A, Huchon G (2006) A review of guidelines for antibacterial use in acute exacerbations of chronic bronchitis. Pulm Pharmacol Ther 19:361-369
41. Obenauer S, Dullin C, Alves F, Missbach-Guentner J, Grabbe E, Heuser M (2007) Flat-panel-detector-based volumetric CT: performance evaluation of imaging for skeletal structures of small animals in comparison to multislice CT. Clin Imaging 31:18-22
42. Mannino DM, Homa DM, Pertowski CA, et al. (1998) Surveillance for asthma--United States, 1960-1995. MMWR CDC surveillance summaries : Morbidity and mortality weekly report CDC surveillance summaries / Centers for Disease Control 47:1-27

43. Nationaal Kompas Volksgezondheid [National Public Health Compass]. http://www.nationaalkompas.nl/gezondheid-en-ziekte/ziekten-en-aandoeningen/ademhalingswegen/astma/omvang. Assessed on June 1, 2013

Summary

Part 1 Introduction

Lung cancer is the most common cause of cancer-related death in the world. The Dutch-Belgian Randomized Lung Cancer Screening Trial (Dutch acronym: NELSON) was launched in 2003, to investigate whether screening for lung cancer by low-dose multi-detector computed tomography (CT) in high risk subjects will lead to a decrease of lung cancer mortality. The NELSON trial is the first lung cancer screening trial, in which lung nodule management is primarily based on nodule volume at first detection, and volume doubling time on follow-up examinations. Nodule volume and volume doubling time are quantitative imaging biomarkers.

Aging and heavy smoking, two participant recruitment criteria in the NELSON trial, are also major risk factors for cardiovascular disease and chronic obstructive pulmonary disease (COPD). Cardiovascular disease is the leading cause of all-cause mortality, causing 32.8% of the total death count in 2008, whereas COPD is the fourth, causing 5.8% of mortality cases. Apart from lung cancer biomarkers, lung cancer screening CT scans yield additional information: coronary calcium scoring and emphysema quantification, as biomarkers of cardiovascular disease and COPD, respectively, can also be performed.

In **Chapter 2**, a practical approach to the radiological evaluation of CT lung cancer screening examinations is described, including assessment of pulmonary nodules and non-nodular diseases. The high resolution images acquired in low-dose thin-slice CT can be used not only to evaluate nodule volume, but also to measure and interpret other thoracic structures, e.g., coronary calcium scoring, emphysema quantification and airway wall thickness.

Part 2 Lung nodules

In lung cancer screening, it is very important to distinct malignant nodules from the vast majority of benign lesions. Prerequisites for accurate distinction are sensitive observer-detection of the pulmonary nodules that are present, and accurate evaluation of nodule growth during follow-up. As described in **Chapter 3**, an anthropomorphic thoracic phantom study was performed to assess the observer detection sensitivity of randomly placed spherical pulmonary nodules, and to determine the accuracy of nodule volumetry. This study shows that a lung cancer screening protocol by low-dose CT such as utilized in the NELSON trial is highly reliable for the detection of spherical pulmonary nodules of 5 mm in diameter (65 mm^3 in volume) and larger. Low-dose CT yields more accurate nodule volumetry when using a semi-automated software tool than in case of manual measurements, with negligible underestimation of actual size.

In addition to spherical pulmonary nodules, small irregular nodules are a common finding in lung cancer screening. The latter nodules have an increased likelihood of being lung cancer. We made a set of artificial spiculated and lobulated nodules to simulate irregular pulmonary nodules. In **Chapter 4**, the observer detection sensitivity and volumetry accuracy were further analyzed using these artificial nodules. This study shows that small irregular solid pulmonary nodules with an actual volume of at least 69 mm^3 can be reliably detected, independent of the observer, CT system and nodule characteristics. However, the volume underestimation of small irregular nodules is larger than in case of spherical nodules. Thus, a lower measured volume threshold for irregular nodules may be needed, as the measured size of irregular nodules can be larger than expected. Also, there is a potential risk to misjudge the growth rate of a small nodule, when a nodule changes into another shape during follow-up.

Repeated CT measurement of the volume of pulmonary nodules can yield different results. To differentiate real growth from measurement variability, a volume increase of at least 25% has been defined as the minimal change to determine nodule growth. In the current practice of lung cancer screening trials, multiple CT systems are involved, often from different vendors. In **Chapter 5**, we assessed the inter- and intra-scanner variability of volumetry for artificial spherical pulmonary nodules. The volume of nodules between 5 and 12 mm diameter correspond well for different 64-multidetector CT scanners in low-dose setting, and when measured repeatedly. The inter- and intra-scanner variability is lower at larger nodule size. Specifically, variability decreases to a maximum of about 5% for ≥ 8 mm nodules. Our results suggest that the commonly accepted cut-off of 25% to determine nodule growth can be reduced for ≥ 8 mm nodules when using semi-automated volumetry. This offers potential to reduce the interval for repeated CT scans in lung cancer screening. An important conclusion from our phantom studies is that semi-automated volume measurements should be used in the setting of lung cancer screening, to obtain accurate volumetry.

Part 3 Coronary artery calcification

The amount of coronary calcification can be evaluated by a dedicated, electrocardiography (ECG)-triggered CT scan of the heart. The extent of coronary calcification, expressed as a coronary calcium score, is a strong predictor of cardiovascular risk. Nowadays, many CTs of the thorax are performed, such as in lung cancer screening. However, these latter scans do not involve ECG-triggering; this can cause motion artifacts of the coronary arteries and may affect the resulting coronary calcium score. Deriving the calcium score from the same examination as used in lung cancer screening may positively impact the cost-effectiveness of screening. In **Chapter 6**, we performed a systematic review and meta-analysis to determine the correlation in calcium scoring between nontriggered and ECG-triggered CT,

and to evaluate the prognostic performance of the calcium score derived from nontriggered CT. There is strong agreement in categorization of calcium scores between ECG-triggered and nontriggered CT techniques. Increasing calcium score categories, based on nontriggered CT, are associated with increasing risk of cardiovascular death or events.

However, variability in coronary calcium quantification based on nontriggered CT is known to be considerable, mainly caused by coronary artery motion. We aimed to determine the influence of motion on the correlation between calcium scores obtained with nontriggered and ECG-triggered CT. For this purpose, an in-depth validation study regarding the impact of motion artifacts on calcium scoring was performed. In **Chapter 7**, the in-vitro agreement in coronary calcium scoring between these two CT techniques was evaluated. We used an anthropomorphic thorax phantom, containing calcifications of known size and density that moved at a controlled velocity. In general, the agreement in calcium scores between nontriggered and ECG-triggered CT is good, but not for large-size and high-density calcifications at higher coronary velocity. Nontriggered CT shows a lower sensitivity to detect small calcifications (calcium score < 25) than ECG-triggered CT.

The results from our studies suggest that calcium scoring based on lung cancer screening CT examinations has potential to stratify individuals into broad calcium score categories, as indicator of cardiovascular risk. However, a zero-score in a screening setting does not reliably exclude presence of coronary calcification.

Part 4 Emphysema and airway wall

The pathogenesis of airflow limitation in COPD is mainly related to emphysema and small airway remodeling. The assumption is that emphysema and peripheral airway wall thickness, as detected by CT, are related to airflow obstruction in COPD patients. Previous studies have shown variable and sometimes conflicting results, and some individual studies have been small and underpowered. In **Chapter 8**, we performed a systematic review and meta-analysis to determine the correlation between CT measurements of emphysema or peripheral airways and airflow obstruction as assessed by lung function tests in COPD. CT measurements of emphysema or peripheral airway are indeed correlated with airflow obstruction parameters in COPD patients. Correlations were found for inspiratory and expiratory CT examinations. These results confirm a relationship between morphology and function in COPD patients. However, this finding is based on multiple measurements in distribution level. It is hard to use this finding on one-to-one relationship between CT parameters and lung function test results. Even so, we can conclude that CT provides a quantitative morphological method to investigate airflow obstruction by emphysema and peripheral airway disease in COPD.

Over 40% of smokers develop chronic bronchitis (CB), a major risk for COPD and all-cause mortality. CB is usually characterized by airway wall remodeling, commonly in large

airways. The morphological changes of airway wall are important to understand the pathogenesis and effect of therapeutic interventions in CB. CT yields quantitative biomarkers of CB, namely different measures of airway wall thickening. In **Chapter 9**, we compare the airway wall thickening along the respiratory pathway between subjects with and without respiratory symptoms of CB, adjusted for relevant confounding factors. The airway wall thickness was higher for the symptomatic than for the asymptomatic group. After adjusting for confounders, respiratory symptoms were related to CT-derived airway wall measurements in airways with diameter of 5 to 10 mm, but not in airways of ≥ 10 mm. From this, we can conclude that airways of 5 to 10 mm diameter optimally reflect the presence of respiratory symptoms in CB, but not the larger airways.

Conclusions

Pulmonary nodule volumetry, coronary calcium scoring, and emphysema and bronchial wall quantification are CT-derived biomarkers for three major diseases: lung cancer, coronary artery disease and COPD, respectively. Firstly, this thesis shows that pulmonary nodule > 5 mm diameter can be reliably detected using low-dose CT. Semi-automated volumetry is more accurate and reproducible than manual assessment. The volumetry variability is lower at larger nodule size. These findings improve the nodule management protocol in lung cancer screening. Secondly, increasing calcium score categories, based on nontriggered CT, are associated with increasing risk of cardiovascular events. A high-risk score detected on nontriggered CT is fairly reliable; however, a zero-score does not reliably exclude the presence of coronary calcification. Thirdly, CT evaluation of emphysema or peripheral airway correlates with airflow obstruction parameters. CT measurement of large bronchial wall associates with respiratory symptoms. These findings confirm a morphological basis of function or symptoms.

As a conclusion, an integrated evaluation of these CT biomarkers may enhance the advantages of CT screening, thus improve evaluation of these three major diseases.

Samenvatting

Deel 1 Evaluatie van longkanker screening CT scans

Longkanker is de meest voorkomende oorzaak van overlijden aan kanker ter wereld. Het Nederlandse Leuvens Longkanker Screenings Onderzoek (NELSON) is gestart in 2003. Deze studie onderzoekt of screenen op longkanker in hoogrisico groepen leidt tot verlaging van de kankersterfte. Screening wordt verricht door middel van multi-detector computer tomografie (CT) met lage stralingsdosis. De NELSON studie is het eerste longkanker screenings onderzoek dat nodulen beoordeelt aan de hand van nodule volume bij eerste detectie, en van volume verdubbelingstijd op vervolgonderzoeken.

Voorwaarde voor deelname aan de NELSON studie zijn hogere leeftijd en fors roken. Dit zijn ook belangrijke risicofactoren voor cardiovasculaire ziekte en chronisch obstructieve longziekte (COPD). Cardiovasculaire ziekte is de meest voorkomende doodsoorzaak, terwijl COPD de vierde doodsoorzaak is. Op het CT onderzoek kunnen afwijkingen gedetecteerd worden die kunnen wijzen op longkanker, genaamd longnodulen. Maar ook andere afwijkingen kunnen beoordeeld worden. Zo kunnen kransslagaderverkalkingen gemeten worden, een maat voor het risico op cardiovasculaire ziekte. Ook kunnen emfyseem en wandverdikking van de luchtwegen geëvalueerd worden, die zijn gerelateerd aan de aanwezigheid van COPD. In **hoofdstuk 2** wordt de aanpak van de radioloog bij de beoordeling van de longkanker screening CT scans beschreven. De hoge resolutie CT beelden, verkregen bij de NELSON studie, worden niet alleen gebruikt om de aanwezigheid van bovenbeschreven afwijkingen te beoordelen, maar ook om deze te meten. Zo wordt het volume van longnodulen berekend middels software, en wordt de hoeveelheid kransslagaderverkalking bepaald en uitgedrukt in een kalkscore. Ook de ernst van emfyseem en wandverdikking van de luchtwegen wordt kwantitatief beoordeeld. Al deze metingen zijn zogenaamde kwantitatieve imaging biomarkers.

Deel 2 Longnodulen

In longkanker screening is het belangrijk dat kwaadaardige nodulen (kanker) worden onderscheiden van overgrote meerderheid van goedaardige longafwijkingen. Een eerste vereiste hiervoor is dat de beoordelaars de aanwezige longnodulen accuraat detecteren. Daarnaast is nauwkeurige evaluatie van groei van longnodulen tijdens follow-up nodig. In de studie beschreven in **hoofdstuk 3** wordt een thorax fantoom gebruikt waarin artificiële ronde longnoduli zijn verspreid. In dit validatie onderzoek is de gevoeligheid van het detecteren van longnodulen bestudeerd, en is de nauwkeurigheid van de volume meting van de nodulen beoordeeld. Deze studie laat zien dat het NELSON screenings protocol waarbij gebruik wordt gemaakt van een lage dosis CT, een hoge betrouwbaarheid heeft voor het detecteren van ronde longnodulen van 5 mm diameter (65 mm^3 in volume) en groter. De

accuratesse in nodule volume metingen is hoger als gebruik wordt gemaakt van semi-automatische software in plaatst van handmatige metingen. Zo wordt het onderschatten van de werkelijke grootte verwaarloosbaar klein.

Naast ronde nodulen worden op longkanker screenings scans ook vaak onregelmatig begrensde longnodulen gevonden, b.v. sprieterige nodulen. De kans op longkanker in deze nodulen is verhoogd. In **hoofdstuk 4** is de gevoeligheid voor detectie, en accuratesse van volume meting van dit type nodulen onderzocht. Hetzelfde thorax fantoom is gebruikt als in **hoofdstuk 2**. Maar nu zijn met de hand gemaakte, onregelmatige nodulen in het fantoom geplaatst. Deze studie laat zien dat kleine onregelmatig begrensde longnodulen met een volume van tenminste 69 mm^3 betrouwbaar kunnen worden gedetecteerd. De onderschatting van het volume van onregelmatige nodulen is wel aanzienlijk groter in vergelijking met ronde longnodulen. Misschien dat in de diagnostische strategie daarom een lager volume afkappunt nodig is voor onregelmatig begrensde nodulen. Daarnaast moet het risico op foutieve berekening van de groeisnelheid in het achterhoofd gehouden worden wanneer kleine longnodulen van vorm veranderen op vervolg scans.

Op herhaal scans verricht met verwaarloosbaar tijdsinterval kan de volume meting van longnodulen variëren. Om groei te onderscheiden van meetvariatie, is een volumetoename van tenminste 25% gedefinieerd. Vanaf 25% volumetoename wordt uitgegaan van groei van de longnodule. In longkanker screening worden in het algemeen verschillende multi-detector CT scanners gebruikt. Dit zou variatie in volume meting tot gevolg kunnen hebben. Daarom is in **hoofdstuk 5** de variabiliteit van volume metingen tussen verschillende scanners en voor herhaald scannen met dezelfde scanner onderzocht. Hiervoor werd hetzelfde thorax fantoom gebruikt met hierin de artificiële ronde longnodulen. Volume metingen van nodulen met diameter tussen 5 en 12 mm blijken goed overeen te komen op lage dosis scans van verschillende 64-multi-detector CT scanners. De variabiliteit vermindert bij grotere nodulen, vooral voor longnodulen met een afmeting van tenminste 8 mm. De studie toont dat het geaccepteerde afkappunt van 25% als definitie voor nodule groei mogelijk verlaagd kan worden voor nodulen met een diameter van 8 mm en meer. Dat zou kunnen betekenen dat het tijdsinterval tot een herhaal CT scan in longkanker screening, nu 3-4 maanden, verlaagd kan worden. De conclusie van de fantoom onderzoeken is dat voor betrouwbare volume metingen in longkanker screening semi-automatische software benodigd is.

Deel 3 Kransslagaderverkalkingen

De hoeveelheid kransslagaderverkalking, uitgedrukt in de kalkscore, is een sterke voorspeller van het risico op cardiovasculaire ziekte. De kalkscore kan accuraat beoordeeld worden op een CT scan van het hart, gemaakt met elektrocardiografische (ECG) triggering. ECG triggering is nodig om beelden te verkrijgen op het moment in de hartcyclus waarop het

hart het minste beweegt, zodat de beelden niet bewogen zijn. Net zoals bij de klinische CT scan van de thorax, wordt bij de scans in het kader van longkanker screening geen ECG triggering toegepast. Hierdoor kunnen de kransslagaders op de beelden bewegingsartefacten vertonen, en dit kan de meting van de hoeveelheid verkalking beïnvloeden. Als de kalkscore bepaald zou kunnen worden op CT scans verricht in het kader van longkanker screening, zou dit een positieve impact kunnen hebben op de kosten-effectiviteit van screening. Dan zou je namelijk met één scan zowel het risico op longkanker als op cardiovasculaire ziekte kunnen bepalen. In **hoofdstuk 6** wordt een systematische review en meta-analyse beschreven naar de accuratesse van kalkscore bepaling voor niet-getriggerde CT scans van de thorax. Hiervoor zijn de kalkscores in categorieën ingedeeld. Kalkscore categorieën gebaseerd op een CT scan van de thorax en die gebaseerd op een ECG-getriggerde CT scan van het hart tonen sterke overeenkomst voor de onderzoeksgroepen als geheel. Vervolgens is gekeken naar de voorspellende waarde van de kalkscore voor later optreden van cardiovasculaire ziekte. Het risico op hart- en vaatziekten is verhoogd in hogere kalkscore categorieën, ook als de kalkscore op niet-getriggerde CT scans gebaseerd is.

Het is bekend dat er aanzienlijke variatie in de meting van de hoeveelheid kransslagaderverkalking is, wanneer deze wordt bepaald op een niet-getriggerde CT scan van de thorax. Meestal wordt dit veroorzaakt door de beweging van kransslagaders, waardoor de beelden onscherp zijn. Om de invloed van bewegingsartefacten op de kalkscore uitkomst van een niet-getriggerde CT scan te bepalen is validatie onderzoek uitgevoerd, zoals beschreven in **hoofdstuk 7**. Middels een fantoom met verkalkingen, bewegend met toenemende snelheid, is de overeenkomst in kalkscore tussen niet-getriggerde CT en ECG-getriggerde CT geëvalueerd. De twee CT technieken tonen een goede overeenkomst in kalkscore, behalve voor grote en zeer dense verkalkingen bij hogere snelheid. Ook is een niet-getriggerde CT scan minder gevoelig voor de detectie van kleine verkalkingen vergeleken met een ECG-getriggerde CT scan. De studies in dit proefschrift ondersteunen de bepaling van de hoeveelheid kransslagader verkalking in longkanker screening trials voor indeling in kalkscore categorieën op populatie niveau. Echter een kalkscore van nul op een niet-getriggerde CT kan de aanwezigheid van kransslagader verkalking niet betrouwbaar uitsluiten.

Deel 4 Emfyseem en wanddikte van de luchtwegen

Emfyseem en veranderingen van kleine luchtwegen spelen een belangrijke rol in het ontstaan van luchtwegobstructie in COPD. Maten van emfyseem en wandverdikking van de kleine luchtwegen gebaseerd op CT scans worden verondersteld samen te hangen met de mate van luchtwegobstructie in COPD patiënten. Eerdere onderzoeken op dit gebied hebben verschillende en soms tegenstrijdige resultaten gevonden. Ook waren sommige studies erg klein. **Hoofdstuk 8** beschrijft een systematische review en meta-analyse, waarin de

relatie wordt bepaald tussen CT kwantificatie van emfyseem en perifere luchtwegen enerzijds, en luchtwegobstructie in longfunctie onderzoek anderzijds. CT metingen van emfyseem en perifere luchtwegen hangen inderdaad samen met luchtwegobstructie gemeten middels longfunctie testen in COPD patiënten. Deze resultaten bevestigen de correlatie tussen morfologie en functie in COPD patiënten. Echter in principe betreft dit een verband dat gevonden is op populatie niveau. Deze bevindingen kunnen niet 1-op-1 toegepast worden op individuele patiënten. Hoe dan ook, geconcludeerd kan wel worden dat CT een kwantitatieve morfologische methode oplevert om luchtwegobstructie bij emfyseem en perifere luchtwegziekten in de COPD populatie te onderzoeken.

Meer dan 40% van de rokers ontwikkelt chronische bronchitis, een risicofactor voor ontwikkeling van COPD en overlijden. Chronische bronchitis wordt in het algemeen gekarakteriseerd door verandering van de luchtwegwand, meestal van de grote luchtwegen. Begrip van de morfologische veranderingen van de luchtwegwand is belangrijk om de oorzaak en het effect van therapeutische interventies in chronische bronchitis te begrijpen. In **hoofdstuk 9** is de wandverdikking van luchtwegen beoordeeld op CT vergeleken tussen personen met en zonder luchtwegsymptomen van chronische bronchitis. De aanwezigheid van luchtwegsymptomen hangt samen met de wanddikte van luchtwegen met diameter van 5 tot 10 mm, na correctie voor beïnvloedende factoren. Een soortgelijk verband werd niet gevonden voor grotere luchtwegen, met diameter van tenminste 10 mm. Hieruit kan afgeleid worden dat m.n. kleinere luchtwegen van 5-10 mm diameter de aanwezigheid van luchtwegsymptomen weerspiegelen, maar niet de grotere luchtwegen.

Conclusie

Volume meting van longnodulen, kalkscore bepaling, en evaluatie van emfyseem en wanddikte van perifere luchtwegen zijn kwantitatieve CT biomarkers voor de drie belangrijkste doodsoorzaken: longkanker, hart- en vaatziekten en COPD. Dit proefschrift laat ten eerste zien dat longnodulen met een diameter van 5 mm en meer betrouwbaar kunnen worden gedetecteerd bij longkanker screening met lage dosis CT. Semi-automatische evaluatie van nodule volume is accurater en beter reproduceerbaar dan meting met de hand, voor zowel ronde als onregelmatige nodulen. De variabiliteit in volume meting neemt af met de nodule grootte, tot onder de afkapwaarde die op dit moment gebruikt wordt om groei van meetvariatie te onderscheiden. Deze bevindingen kunnen helpen om de management van bij CT screening gedetecteerde longnodulen te optimalizeren. Ten tweede wordt geconcludeerd dat toenemende hoeveelheid kransslagader verkalking, gebaseerd op niet-getriggerde CT van de thorax, gerelateerd is aan verhoogd risico op hart- en vaatziekten op populatie niveau. Een hoge kalkscore, gedetecteerd op een screenings CT, is redelijk betrouwbaar; echter een kalkscore van nul sluit de aanwezigheid van kransslagaderverkalking niet uit. Ten derde, CT evaluatie van emfyseem correleert met resultaten van longfunctie

onderzoek op populatie niveau. Wanddikte meting van perifere luchtwegen op basis van CT hangt samen met de aanwezigheid van luchtweg symptomen in rokers. Deze bevindingen bevestigen een morfologische basis van longfunctie en symptomen.

In conclusie, longkanker screening CT kan niet alleen aangewend worden voor evaluatie van longnodulen, maar levert ook de mate van kransslagaderverkalking, emfyseem en wanddikte van perifere luchtwegen op. Geïntegreerde evaluatie van deze CT biomarkers voor de drie grote doodsoorzaken in de westerse bevolking (longkanker, cardiovasculaire ziekte en COPD) kan een positieve impact hebben op de inzet van CT screening.

Acknowledgements

At the moment that I so closely approach my PhD title, I would like to express my deep and sincere gratitude to all of those who, in different ways, have walked beside me, offering encouragements, supports and cooperations in this process.

Foremost, I would like to express my sincere gratitude to my promoter, Professor Matthijs Oudkerk for his continuous and valuable supports. I enjoyed working in his researching team. I highly appreciated his professional guidance, consistent support and generous encouragement. Great thanks for his innovative ideas on my projects. Also great thanks for offering me the opportunity to enrol as a PhD candidate and to push me forward with his scientific insights and creative thinking.

My sincere gratitude also goes to my co-promotor, Dr. Rozemarijn Vliegenthart for her great efforts contributed towards the completion of this thesis. Great thanks for her detailed comments and careful consideration of my writings. Also great thanks for the research coordination, scientific insight, enormous contact, valuable discussion and exhaustive revision of all my manuscripts. Without her help, I could not have arrived at the final stage of my thesis.

Dear Dr. Marcel Greuter, thank you for your continuous guidance and supervision, which broadened my knowledge in medical physics. Professor Truuske de Bock, thank you for your heuristic guidance, which sparked my thinking in statistics. Dr. Pim de Jong, thank you for the help to perform phantom studies, and your valuable comments on many of my manuscripts. Dr. Stella Noach, it is really my pleasure to work with you. Thank you for being there to answer my questions and support my work.

Professor Harry Groen, thank you for the scientific guidance and coordination on the airway study and Akkelies Dijkstra for the help to perform this study. I really enjoyed working with you. Dr. Judith Vonk, thank you for detailed guidance on statistics in this study. And Dr. Nick ten Hacken, thank you for your professional comments relevant to the knowledge on respiratory medicine.

My dear friends: Zhao Yingru, Wang Hao, Pieter de Jong, Wang Ying, Zhou Lu, Wang Rong, Lv Congchao, Qian Cheng, Wang Hongwei and other good friends, I will remember the joyful dinners, chatting and other relaxed times together and cherish our friendship in my life.

My colleagues: Astri Handayani, Gert Jan Pelgrim, Gonda de Jonge, Hilderbrand Dijkstra, Jaap Groen, Kadek Aryanto, Karolien Jaspers, Marjolein Heuvelmans, Martijn den Dekker, Monique Dorrius, Pandji Triadyaksa, Peter van Ooijen, Petra Serbanescu Kele, Volkan Tuncay, Wisnumurti Kristanto, and other colleagues in Groningen: I really enjoyed the great times of cooperation we had. Wim Tukker and Jamal Moumni, thank you for doing overtime jobs to scan the phantoms. Annick Scheeren, Cevahir Sahbaz, Marc Marieke

and Vera Overkempe, thank you for working together on the Lungman study. Martin Willemink, thank you for your coming to evaluate the pulmonary nodules from Utrecht.

I would also like to thank the members of the reviewing committee for this thesis, Professor Jan-Willem Lammers, Professor Willem Mali and Professor Harry Groen for the critical reading of my manuscripts.

Dear Professor Miao Jingtao, thank you for recommending me to Groningen for a PhD study. You inspired me a lot, not only in scientific area, but also your wisdom in the attitude of life. Professor Guo Qiyong, my boss of my master degree, thank you for leading me into the beautiful scientific world. I also want to thank Professor Zhang Guixiang and Professor He Zhiyan in Shanghai, who inspired me in diagnostic radiology.

Finally, I would like to thank my family for their consistent support and selfless care on my life. Mum and dad, thank you for what you have done since I was born. Yu, my beloved wife, I enjoyed the journey we have passed in our life. Thank you for the endeavour to raise our baby. Weiwei, my lovely daughter, I like the new role you brought to me as a daddy.

About the author

Xueqian Xie was born on February 6, 1977 in Fushun, China. After graduation from Fushun No.2 Senior Middle School in 1996, he subsequently started his education in clinical medicine at the China Medical University in Shenyang, China. In 2001, he obtained his medical degree and started the clinical training in diagnostic radiology. Simultaneously, he performed scientific research on the development of hepatic CT perfusion software for his master degree, supervised by Prof. Guo Qiyong at the China Medical University Affiliated Shengjing Hospital. After graduation from university in 2003, he worked as a radiologist at the Shanghai Jiaotong University Affiliated First People Hospital in Shanghai.

In December 2009, he came to the Netherlands to start his PhD project entitled 'CT Biomarkers in lung cancer screening' at the University of Groningen, Center for Medical Imaging – North East Netherlands (CMI^{NEN}), Department of Radiology, University Medical Center Groningen, The Netherlands, supervised by Prof. Matthijs Oudkerk and Dr. Rozemarijn Vliegenthart. The results of the PhD research are presented in this thesis.

List of Publications

1. **Xie XQ**, Greuter MJW, Groen JM, de Bock GH, Oudkerk M, de Jong PA, Vliegenthart R. Can nontriggered thoracic CT be used for coronary artery calcium scoring? A phantom study. Medical Physics. 2013 Jul 18. [Epub ahead of print]
2. **Xie XQ**, Willemink MJ, de Jong PA, van Ooijen PMA, Oudkerk M, Vliegenthart R, Greuter MJW. Small irregular pulmonary nodules on low-dose CT: observer detection sensitivity and volumetry accuracy. (Accepted by American Journal of Roentgenology)
3. **Xie XQ**, Heuvelmans MA, van Ooijen PMA, et al. A practical approach to radiological evaluation of CT lung cancer screening examinations. (Accepted by Cancer Imaging)
4. **Xie XQ**, Willemink MJ, Zhao YR, et al. Inter-scanner and inter-observer variability of pulmonary nodule volumetry on low-dose 64-row CT: an anthropomorphic phantom study. British Journal of Radiology. 2013 Jul 24. [Epub ahead of print]
5. **Xie XQ**, Zhao YR, de Bock GH, de Jong PA, Mali WP, Oudkerk M, Vliegenthart R. Validation and prognosis of coronary calcium scoring in nontriggered CT: systematic review and meta-analysis. Circulation: Cardiovascular Imaging. *2013;6(4):514-521.*
6. Horeweg N, van der Aalst CM, Vliegenthart R, Zhao YR, **Xie XQ**, Scholten ET, Mali WP, Thunnissen FB, Weenink C, Groen HJM, Lammers JWJ, Nackaerts, K, van Rosmalen J, Oudkerk M, de Koning, HJ. Volumetric computer tomography screening for lung cancer: three rounds of the NELSON trial. European Respiratory Journal. 2013 Jul 11. [Epub ahead of print]
7. Heuvelmans MA, Oudkerk M, de Bock GH, de Koning HJ, **Xie XQ**, van Ooijen PMA, Greuter MJW, de Jong PA, Groen HJM, Vliegenthart R. Optimisation of volume-doubling time cut-off for fast-growing lung nodules in CT lung cancer screening reduces false-positive referrals. European Radiology. 2013, 23(7):1836-1845.
8. **Xie XQ**, Zhao YR, Snijder RA, van Ooijen PMA, de Jong PA, Oudkerk M, de Bock GH, Vliegenthart R, Greuter MJW. Sensitivity and accuracy of volumetry of pulmonary nodules on low-dose 16- and 64-row multi-detector CT: an anthropomorphic phantom study. European Radiology. 2013, 23(1):139-147.
9. **Xie XQ**, de Jong PA, Oudkerk M, Wang Y, Ten Hacken NH, Miao J, Zhang G, de Bock GH, Vliegenthart R. Morphological measurements in computed tomography correlate with airflow obstruction in chronic obstructive pulmonary disease: systematic review and meta-analysis. European Radiology. 2012, 22(10):2085-2093.
10. Wang H, Zheng LF, Feng Y, **Xie XQ**, Yang XM, Zhang GX. CTA combined with CT perfusion for assessing the efficacy of anti-angiogenic therapy in rabbit VX2 tumors. Academic Radiology. 2012, 19(3):358-365.

11. Wang H, Zheng LF, Feng Y, **Xie XQ**, Zhao JL, Wang XF, Zhang GX. A comparison of 3D-CTA and 4D-CE-MRA for the dynamic monitoring of angiogenesis in a rabbit VX2 tumor. European Journal of Radiology. 2012, 81(1):104-110.
12. Zhao YR, **Xie XQ**, de Koning HJ, Mali WP, Vliegenthart R, Oudkerk M. NELSON lung cancer screening study. Cancer Imaging. 2011, 11(Spec No A):S79-S84.
13. Ying HY, Wang F, **Xie XQ**. Three-dimensional computed tomography reconstruction of the eyes before and after removing the scleral encircling buckles. Chinese Journal of Ocular Fundus Diseases. 2010, 26(6): 536-539. (In Chinese)
14. Wang H, Guo R, Zhang GX, Shi XY, Zhao JL, Zheng LF, **Xie XQ**. Preliminary research on dendrimer-entrapped gold nanoparticle as a CT molecular probe. Chinese Journal of Biomedical Engineering. 2010, 16(1): 10-14. (In Chinese)
15. Xia SJ, Liu JH, He ZY, Lu J, **Xie XQ**. Assessment the value of 16-slice multi-detector CT in pelvicaliceal system map in MPCNL. Chinese Journal of Urology. 2007, 28(1): 19-22. (In Chinese)
16. Song JH, Miao JT, **Xie XQ**. Multi-slice computed tomography pulmonary perfusion imaging and clinic cardiac function in patients with chronic cardiac insufficiency: a preliminary correlative study. Chinese Journal of Medical Imaging Technology. 2007, 23(11): 1628-1632. (In Chinese)
17. Jin LF, Miao JT, **Xie XQ**, Chen YJ, Lou XJ. The characteristics of pulmonary CT perfusion imaging in the patients with chronic obstructive pulmonary disease. Journal of Practical Radiology. 2007, 23(11):1457-1460. (In Chinese)
18. Jin LF, Miao JT, Li ZY, **Xie XQ**, Dai XM. Study of voxel index from multislice CT in evaluating chronic obstructive pulmonary disease air-trapping. Chinese Journal of Medical Imaging Technology. 2007, 23(7):1033-1036. (In Chinese)
19. Zhang H, Shen TZ, Chen XR, Miao JT, **Xie XQ**. Evaluation of brain metastases with dynamic susceptibility-contrast MR imaging. China Radiology Journal. 2006, 40(9):903-907. (In Chinese)
20. Zhang H, Shen TZ, Chen XR, Miao JT, **Xie XQ**. Value of perfusion MR imaging in the differential diagnosis between single brain metastasis and high grade gliomas. China Radiology Journal. 2006, 40(4):393-396. (In Chinese)
21. Miao JT, Xu HF, **Xie XQ**. Enterprise PACS of group hospital: implement consideration. Chinese Computed Medical Imaging. 2006, 12 (5):321-324. (In Chinese)
22. Wang PW, He ZY, Yuan XC, **Xie XQ**, et al. Clinical and image study of pancreatic hemocirculatory change in patients with acute pancreatitis. Chinese Journal of Digestion, 2005, 25 (8):469-472. (In Chinese)
23. Fan GM, Guo QY, **Xie XQ**, et al. Experiment study about the relationship between contrast media injection dose and flow rate in CT perfusion imaging. Chinese Journal of Medical Imaging Technology, 2005, 21(1):14-17. (In Chinese)

24. **Xie XQ**, Guo QY, Jiang J. Development of hepatic CT perfusion software on personal computer. China Radiology Journal, 2004, 38(12): 1265-1268. (In Chinese)
25. Miao JT, Xu HF, Zhang JX, Zhang LF, Miao WX, **Xie XQ**, et al. Cost/effectiveness of digital radiology department: evaluation and management. Chinese Journal of Hospital Administration. 2004, 20 (6):359-360. (In Chinese)
26. Guo Q, Hao F, Duan X, **Xie XQ**, Liao W. Multi personal computer storage system: solution of sea capacity PACS storage. Chinese Medical Journal (English version). 2003, 116(5):650-653.
27. **Xie XQ**, Bai SL. Genetic regulation and control of hypertrophic cardiomyopathy. Progress of Anatomical Sciences. 2001, 1(7):57-60. (In Chinese)